Fall 2013

# A Novel Multistage Image Registration Technique with Graph-based Region Descriptors

Francis Ryals Bowen
*Purdue University*

Graduate School ETD Form 9
(Revised 12/07)

# PURDUE UNIVERSITY
## GRADUATE SCHOOL
## Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Francis Bowen

Entitled
A Novel Multistage Image Registration Technique with Graph-based Region Descriptors

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Eliza Du, Co-Chair
Chair

Xukai Zou

Jianghai Hu , Co-Chair

Johnny Park

Lingxi Li

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): Jianghai Hu , Co-Chair

Approved by: M. R. Melloch 11/20/13
Head of the Graduate Program Date

# A NOVEL MULTISTAGE IMAGE REGISTRATION TECHNIQUE WITH GRAPH-BASED REGION DESCRIPTORS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Francis Bowen

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2013

Purdue University

West Lafayette, Indiana

This thesis is dedicated to my honus.

# ACKNOWLEDGMENTS

I would like to offer my sincerest appreciation to my advisors, Dr. Eliza Y. Du and Dr. Jianghai Hu, for their immense support and guidance during this doctoral work. Their contributions and advice have helped influence and shape a body of work that I am truly proud of. The invaluable feedback received has made me a better researcher and will forever inspire my career decisions.

I would also like to express my gratitude to my Ph.D. committee that includes Dr. Johnny Park, Dr. Lingxi Li, and Dr. Xukai Zou. Their feedback has provided substantial improvement to the overall idea and presentation of this work.

Additionally, I would like to thank the faculty and staff of the Electrical and Computer Engineering Department at Indiana University-Purdue University Indianapolis. The faculty at IUPUI, specifically Dr. Maher Rizkalla, has provided opportunities that I could not have found elsewhere. I also would like to express my appreciation to Mr. Matt Golden and Rob Meagher for the assistance they provided along every step of my journey.

Lastly, I would like to thank my family; Jennifer and Ryals, parents, Rosario and Peter, step-dad, Mark, brother, Charles, sister, Stacey, and father-in-law, Jim, who have always provided me with constant support and inspiration. None of this work would have been possible without their backing. I will always be grateful for my family's ability to take my goals and make them their own through support and motivation.

TABLE OF CONTENTS

Page

LIST OF TABLES ..... vi
LIST OF FIGURES ..... vii
ABSTRACT ..... xi
1. INTRODUCTION ..... 1
1.1. Background ..... 1
1.2. Motivation ..... 3
1.3. Challenges ..... 5
1.3.1. Photometric Variations ..... 5
1.3.2. Geometric Variations ..... 6
1.3.3. Invariance to Rotation and Scale ..... 7
1.3.4. Robust Registration with Local Deformations ..... 8
1.4. Thesis Contributions ..... 10
1.5. Outline ..... 11
2. LITERATURE REVIEW ..... 12
2.1. Image Registration ..... 12
2.1.1. Feature-based Registration ..... 13
2.1.2. Area-based Registration ..... 17
2.1.3. Image Similarity Measures ..... 19
2.1.4. Fourier Transform Registration ..... 22
2.2. Image Features ..... 24
2.2.1. Feature Point Descriptors ..... 24
2.3. Region Descriptors ..... 30
2.4. Summary ..... 33
3. GRAPH-BASED INVARIANT REGION DESCRIPTOR ..... 34
3.1. Graph-based Descriptors ..... 36
3.2. Proposed Graph-based Descriptor ..... 36
3.2.1. Clustering ..... 38
3.2.2. Graph Structure ..... 43
3.3. Graph-based Descriptor Matching ..... 46
3.3.1. Spatial and Descriptor Characteristic Matching ..... 46
3.3.2. Angle and Descriptor Characteristic Matching ..... 48
3.4. Results ..... 54
3.4.1. Graph Matching ..... 54
3.4.2. Feature Matching Across Images ..... 55
3.4.3. SIFT, SURF and BRISK Matching Comparisons ..... 58

Page

3.5. Summary....63
4. COMPREHENSIVE FEATURE AND TEXTURE FUSION-BASED IMAGE REGISTRATION APPROACH....64
4.1. Coarse Image Search....65
4.2. Fine Image Search....67
4.3. Transformation Parameter Estimation....70
4.4. Results....71
4.5. Summary....74
5. MULTISTAGE FEATURE-BASED IMAGE REGISTRATION....76
5.1. Coarse Registration....78
5.1.1. Translation Estimation....79
5.1.2. Rotation Estimation....83
5.2. Registration Evaluation....92
5.3. Limited Neighborhood Graph-based Region Descriptor Registration....96
5.4. Feature-based Registration with Coarse Area Search....98
5.5. Graph-based Region Descriptor Registration....100
5.6. Alternate Invariant Feature Points....101
5.6.1. Limited Window Graph-based Registration with BRISK....102
5.6.2. Comprehensive Registration with BRISK....103
5.6.3. Unrestricted Graph-based Registration with BRISK....103
5.7. Summary....104
6. RESULTS AND DISCUSSION....105
6.1. Initial Coarse Registration....105
6.2. Limited Window Graph-based Region Descriptor Registration....113
6.3. SURF-based Registration with Coarse Template Matching....118
6.4. Unrestricted Graph-based Region Descriptor Registration....122
6.5. Overall Results....128
6.5.1. Class 1 Result Comparisons....132
6.5.2. Class 2 Result Comparisons....135
6.5.3. Class 3 Result Comparisons....138
6.6. GPU Implementation Considerations....141
6.7. Summary....142
7. CONCLUSIONS AND FUTURE WORK....144
7.1. Conclusions....144
7.2. Thesis Contributions....147
7.3. Future Work....148
7.4. Publications....149
LIST OF REFERENCES....150
A. TIME ANALYSIS AND PROPOSED GPU IMPLEMENTATION....160
VITA....177

# LIST OF TABLES

Table Page

3.1 Gray Code encoding for base-10 range 0-7. The red indicates the single bit changed for successive value changes ..... 51

3.2 Matching accuracy distribution parameters. ..... 62

6.1 Mutual information scores per class compared to traditional intensity and feature-based approaches ..... 131

6.2 Overall average mutual information or proposed registration method compared to no registration as well as intensity and feature-based registration approaches for entire the test set ..... 132

6.3 Estimated speedup factors on a GPU ..... 142

A.1 Number of computations required for SURF feature point detection and descriptor extraction ..... 164

A.2 Complexity summary of the multistage registration approach ..... 170

A.3 Average execution times for each phase of each stage ..... 171

A.4 Speedups using GPU architecture ..... 175

A.5 Estimated average execution times for the phases of each stage of the proposed method while utilizing a GPU ..... 175

# LIST OF FIGURES

Figure Page

1.1 Example image sets that pose challenges for image registration techniques.....3

1.2 Example image set illustrating photometric variations due to a fire .....6

1.3 Residential structures from different perspectives and lighting conditions.....7

1.4 Example of rotational differences in aerial images.....8

1.5 Clustered invariant feature points where the green dot denote invariant feature points and the clusters are defined as the feature points contained within each red bounding box .....9

1.6 Sub-graph comparisons.....10

2.1 Registration points identified through SURF keypoints and triangulation [27].....14

2.2 Registration result using MCE [34].....18

Fig. 2.3 Polar transformation .....23

2.4 Box filters for Lyy and Lxy.....26

2.5 BRISK sampling pattern for $k = 1$ [50].....29

3.1 Region descriptor overview.....37

3.2 Detected feature point clusters (a), and a close view (b) of the single cluster indicated by the blue outline in (a) .....40

3.3 Example clustered feature points using the k-means approach with a 2D Gaussian. Feature points indicated in the same color belong to the same cluster .....42

3.4 Two example connected graphs from invariant feature points. The green markers indicate the feature points while the blue lines are the connected edges of the graph .....44

3.5 Example SURF descriptors and their detected scales [46].....45

3.6 Sample graphs created using greedy node selection with feature point filtering .....46

Figure Page

3.7 Partial graph matching ..... 48

3.8 8-bit BAM representation ..... 50

3.9 Visual representation of the angle descriptor components ..... 53

3.10 Two examples of matched graphs using SURF feature points and their angles ..... 55

3.11 Example image pairs demonstrating the feature matching abilities of the proposed region descriptor ..... 58

3.12 Graphs illustrating the percent of feature points matched between 41 sets of image pairs ..... 61

3.13 Matching accuracy distribution for the proposed region descriptor, SURF, SIFT and BRISK. ..... 63

4.1 Algorithm overview ..... 65

4.2 Reference image (left) and segmented query image (right) ..... 65

4.3 Similar patches identified using the normalized cross-correlation ..... 66

4.4 Example SURF matching with and without symmetric matching ..... 69

4.5 Control point matching ..... 70

4.6 Matched SURF feature points from a reference (a) and query image (b). The resulting registration utilizing the SURF points and the DLT method for transformation estimation (c) ..... 71

4.7 Reference and query images shown with similar patches denoted in red boxes and identified control points designated with green circles (a). The remapped query image after registration (c) ..... 72

4.8 Registration results depicting the reference images (a-d), query images (e-h), and transformed query image (i-l). The red boxes indicate the matched image patch while the green circles denote the detected feature points ..... 74

5.1 General overview of the tasks involved in the proposed image registration method ..... 76

5.2 Detailed overview of proposed registration system ..... 77

5.3 Proposed Coarse Registration ..... 78

5.4 Cross-correlation matrix showing a peak at (543,269) which indicates a translation of (-58,-22) is required. The XYZ (a), XZ (b) and YZ(c) views are provided for illustrative purposes ..... 82

Figure Page

5.5 Aerial image depicting a residential neighborhood after a fire. Original reference image (left) and translated query image (right) ....82

5.6 Example original (a) and calculated Canny edge maps (b) ....87

5.7 (a) Structuring element SE used by the erosion operator, (b) binary image I, (c) illustration identifying the image elements removed by the operator, and (d) the resulting image after erosion by SE ....88

5.8 (a) Structuring element SD used by the dilation operator, (b) binary image I, (c) illustration identifying the image elements added by the operator, and (d) the resulting image after dilation by SD....89

5.9 (a) Structuring elements, SD and SE, (b) binary image I, (c) dilation of I, (d) erosion of dilated I, and (e) resulting binary image from dilation followed by erosion....90

5.10 (a) Original canny edge map, (b) binary image after closing operation, and (c) binary image after dilating closed image according to MI, SD, SE ....91

5.11 (a) Original image, (b) image edge map after morphological operations, and (c) log-polar edge map....92

5.12 (a-c) R,G, and B histograms for original query image, (d-f) R,G, and B histograms for accurately registered query image ....94

5.13 (a-c) R,G, and B histograms for original query image. (d-f) R,G, and B histograms for inaccurately registered query image ....95

5.14 Proposed limited neighborhood graph-based registration ....96

5.15 Limited search neighborhood where the red square represents the centroid of the reference graph while green dots denote centroids of graphs in the query image that lie within the search window ....97

5.16 Reference image (left) and query image (right) with matched region descriptors ....98

5.17 Proposed intensity and feature-based registration method ....98

5.18 Example matched images patches and identified registration control points ....100

5.19 Overview of proposed registration technique using unrestricted matching of graph-based region descriptors ....100

5.20 Example matches using the unrestricted graph-based region descriptor matching method....101

6.1 RMSE for rotation angles ranging from 1° to 180° ....106

Figure Page

6.2 Average RMSE for all combinations of x and y translation from 1 to 50 pixels .....107

6.3 Coarsely registered images from an aerial view of a natural scene .....109

6.4 Coarsely registered images from an aerial view of structural content .....111

6.5 Failed coarse registration of an urban scene .....112

6.6 Example region descriptor matching results .....115

6.7 An aerial view of structural images with large differences in pixel intensities .....116

6.8 Residential registration example using the limited window region descriptor method .....118

6.9 Aerial image registered using NCC template matching and SURF .....120

6.10 Street view of a registered building using NCC template matching and SURF .....122

6.11 Registration of an aerial view of a natural scene .....124

6.12 Registration of a street view building during a disaster .....125

6.13 Registration of an aerial view of a residential area greatly affected by fire damage .....126

6.14 Registration of a street view building after significant damage has been sustained during a riot .....127

6.15 Mutual information comparisons before and after registration .....130

6.16 Example class 1 registration results comparing the proposed method to common intensity and feature-based methods .....135

6.17 Example class 2 registration results comparing the proposed method to common intensity and feature-based methods .....138

6.18 Example class 3 registration results comparing the proposed method to common intensity and feature-based methods .....141

A.1 Nvidia CUDA hardware architecture .....161

A.2 Time allocations for each stage of the proposed method .....172

A.3 Relationship between the number of detected region descriptor graphs and the number of initial feature points detected .....173

A.4 Number of feature points related to execution time for each proposed stage .....174

# ABSTRACT

Bowen, Francis R. Ph.D., Purdue University, December 2013. A Novel Multistage Image Registration Technique Using Graph-based Region Descriptors. Major Professors: Eliza Y. Du and Jianghai Hu.

Successful image alignment is an essential function for many image processing methods. The geometric and photometric variations between images adversely affect the ability for an algorithm to estimate the transformation parameters that relate the two images. Local deformations, lightning conditions, object obstructions, and perspective differences all contribute to the challenges faced by traditional registration techniques. In this work, a novel multistage registration approach is proposed that is resilient to view point differences, image content variations, and lighting conditions. The proposed method is demonstrated to be effective for registration scenarios involving images of a scene or object before and after a disaster. Robust registration is realized through the utilization of a novel region descriptor which couples the spatial and textural characteristics of invariant feature points. Clusters of invariant feature points are shown to provide more discriminative features than the traditional point descriptors.

The multistage method is a hierarchy of registration approaches that takes advantage of feature, intensity and Fourier-based techniques. The three phases include a limited search window method that employs the proposed graph-based region descriptor, a comprehensive approach which fuses intensity and feature-based analysis, and an exhaustive search method which also utilizes the region descriptor. Each successive stage of the registration technique is evaluated through an effective similarity metric which determines subsequent action. The registration of aerial and street view images from pre and post disaster provide strong evidence that the proposed method estimates

more accurate global transformation parameters than traditional intensity and feature-based methods. Experimental results involving the mutual information metric confirm the robustness and accuracy of the proposed multistage image registration methodology. Moreover, experimental results show that the proposed graph-based region descriptor offers higher matching accuracy than SIFT, SURF and BRISK descriptors for the test set of images from before and after a disaster.

# 1. INTRODUCTION

## 1.1. Background

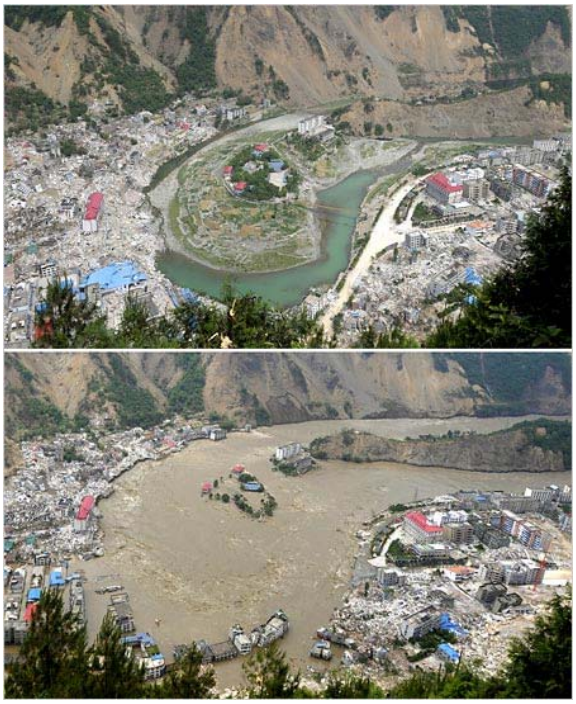
For many image processing tasks, such as image fusion or stitching, image registration is often exploited as a preprocessing step [1-3]. Registration is a necessary task that estimates the transformation parameters relating two images. The purpose is to project one image such that both images contain some region which overlaps and may appear to be from the same perspective. This region of interest isolates areas of the images that may share common details and features.

Registration requirements may vary greatly amongst different applications. For instance, multi-modal medical registration will rely heavily upon shape context, while general object registration will necessitate the analysis of textural properties. Two matched images may vary geometrically through translation, rotation, affine, and perspective transformations, or photometric variations from object occlusion and pixel dissimilarities from lighting differences. Images of a scene before and after a disaster is particularly difficult to register due to many dissimilarities in geometric and photometric composition, however analysis of such images is an important goal for disaster management and risk planning [4,5]. Fig. 1.1a depicts two residential structures from two vastly different perspectives, while Fig. 1.1b and Fig. 1.1c illustrate scenes from similar perspectives but vary immensely in appearance due to a fire and natural disaster, respectively. Both examples provide challenging scenarios for image registration techniques.

Registration algorithms can be broadly classified as feature-based or area-based approaches. In the former, distinct image features such as corners [6], gradient edges [7], or shape descriptors [8] are used to define the geometrical mapping between images. Feature-based approaches rely on the detected feature's invariance to affine, rotational, and translational transformations. With area-based methods, pixel intensities are compared directly for a sub-region of an image. In such a scenario, a similarity measure is coupled with an optimization algorithm in an attempt to identify the closest mapping of pixels [9].


a. Residential structure from different perspectives


b. Residential structure from the same perspective but varying textural properties



c. An aerial image with photometric and geometric differences

Fig. 1.1 Example image sets that pose challenges for image registration techniques

## 1.2. Motivation

Vast amounts of data captured before and after a disaster are often collected for future analysis. Effective processing of such data may include image registration of two images with limited mutual information. The ability to automate this process will lead to faster disaster management and response. More specifically, a multi-phase image registration system with a novel region descriptor is proposed which can be utilized as a preprocessing task for Geographical Information Systems (GIS) platforms. The improvement of the registration stage will positively affect subsequent processing tasks, such as fusion, stitching, or the identification of common areas of interest.

Traditional processing approaches are not always suited for dealing with the large photometric and geometric variations that exist in an image set that compares a scene before and after a disaster. Thomase *et al.* [10] propose registration approaches that employ invariant feature points for disaster image alignment; however image content can vary tremendously where image gradients are skewed or completely different, therefore a particular feature point may not be suitable for many scenarios. In [11] we proposed a graph-based region descriptor for accurately matching features across such image sets. It is the aim of this work to improve image registration in difficult situations through an innovative multi-stage approach that exploits a novel region descriptor as well as area-based analysis for a broad ability to register two images greatly affected by disasters. Such a method will be applicable to aerial and urban imaging where various levels of transformations may invalidate other approaches. Many other applications could benefit from the proposed system where traditional registration methods are inaccurate due to the local geometric and photometric deformations.

We propose a comprehensive 3-phase image registration method that takes advantage of feature point detection but imposes a strict method for identifying optimal interest points for the estimation of the homography matrix. Invariant feature points and their spatial relationships are leveraged to identify the suitable control points for the estimation of the transformation parameters required for accurate registration. A k-nearest neighbor graph constructed from a collation of Speeded-up Robust Feature (SURF) feature points is described along with the matching criteria for the region descriptor and subsequent control point identification. Previous work in [11] demonstrates the effectiveness of the proposed region descriptor for image feature matching while this work extrapolates the core concepts to image registration. Finally, the direct linear transform (DLT) approach is utilized for the transformation parameter estimation. The proposed approach will be tested using a variety of images from disaster scenes in an urban environment as well as aerial photography.

### 1.3. Challenges

In order to achieve the desired registration goals, the proposed system must overcome a set of challenges. The three-phase system attempts to detect these conditions on-the-fly to select the optimal approach.

#### 1.3.1. Photometric Variations

Large differences in pixel intensities or textural properties will adversely affect the ability to register two images as common features may be indistinguishable. Pixel intensities may vary due to differences in lighting conditions, as well as damage caused by natural disasters, such as from a fire or a flood. An example is provided in Fig. 1.2a and Fig. 1.2b, where a building is shown before and during a fire, respectively. Although the viewpoints and lighting conditions are similar, a significant section of the images differ in pixel characteristics due to the fire and resulting smoke.


a. Urban scene before disaster


b. Urban scene during a fire in London 2011

Fig. 1.2 Example image set illustrating photometric variations due to a fire

### 1.3.2. Geometric Variations

When studying images pre and post disaster, geometric variations are often present due to damage. Buildings that sustain large amounts of destruction may be unrecognizable after a disaster. Additionally, perspective differences when acquiring the images also add to the geometric differences between image pairs. Fig. 1.3 illustrates view point changes for the same scene where the geometric differences may affect registration results. As shown, dissimilarities in both scale and perspective are present.

Fig. 1.3 Residential structures from different perspectives and lighting conditions

### 1.3.3. Invariance to Rotation and Scale

The proposed region descriptor should be invariant to rotational and scale variations in order to be robust under the most common geometric transformations. Aerial and satellite imagery is often acquired at different times and will therefore introduce rotational and scale differences. The aerial image provided in Fig. 1.4 depicts a common scenario where images captured from an aircraft or UAV differ only by a small amount of translation and rotation. These images can then be combined through image registration and stitching to give a global view of a particular scene [12,13].

Fig. 1.4 Example of rotational differences in aerial images

### 1.3.4. Robust Registration with Local Deformations

The proposed approach attempts to utilize invariant feature points and their spatial and textural properties to identify a novel region descriptor that will serve as the basis for control point identification. Such an approach poses a unique set of challenges such as the structural properties of the region descriptor as well as feature matching.

#### 1.3.4.1. Structural Properties of a Graph-based Region Descriptor

Careful consideration must be taken in how to structurally form the proposed region descriptor. Several methodologies will be outlined and compared. The fundamental structure of the region descriptors will ultimately affect the invariance to photometric, geometric, rotation, and scale variations. Furthermore the overall structure will be highly dependent on how feature points are grouped, therefore two suitable approaches will be proposed. Fig. 1.5 provides an example image where invariant feature points are extracted, clustered and considered for creating the proposed region descriptor.

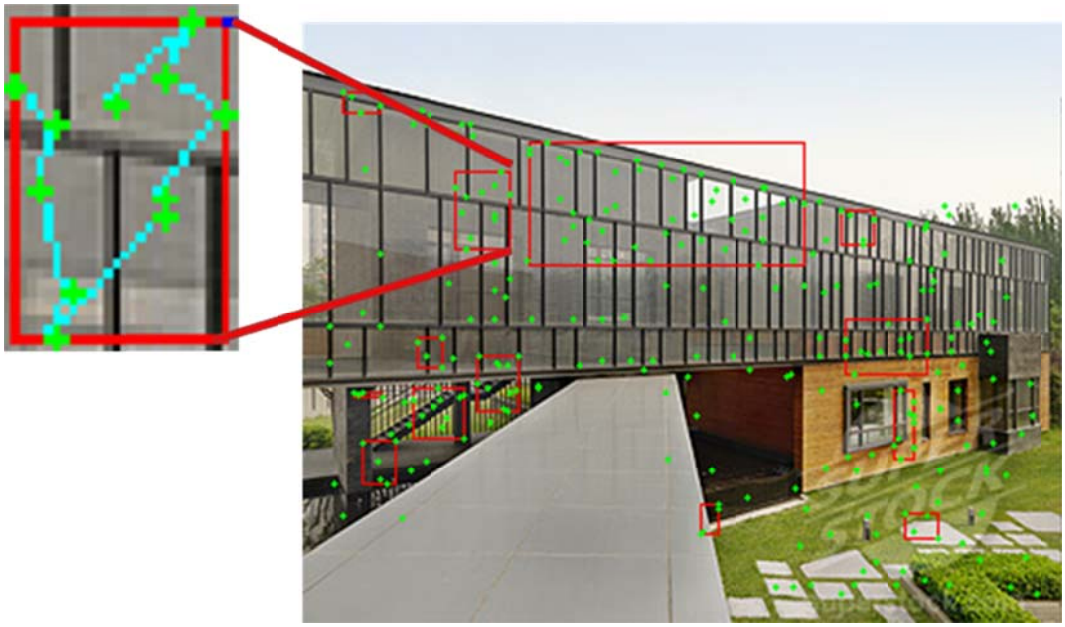


Fig. 1.5 Clustered invariant feature points where the green dot denote invariant feature points and the clusters are defined as the feature points contained within each red bounding box

#### 1.3.4.2. Region Descriptor Matching

An appropriate similarity measure for comparing the proposed region descriptor is crucial to the accuracy of identifying suitable control points for the subsequent image registration. Spatial and descriptor characteristics should be coupled such that the proposed similarity score will result in a large value for matched graphs while mismatched graphs are assigned a small score. Moreover the score should be discriminative enough that there is a one-to-one mapping between region descriptors across image pairs. The comparison of two different sized graphs poses another challenge for graph-based region descriptor matching. Fig. 1.6 illustrates an example comparison where nodes 1, 2, and 3 of $G_1$ are compared to $G_2$. Many options exist for sub-graph comparisons; however an exhaustive search of all sub-graph comparisons is not computationally reasonable, therefore a method for graph matching is proposed.

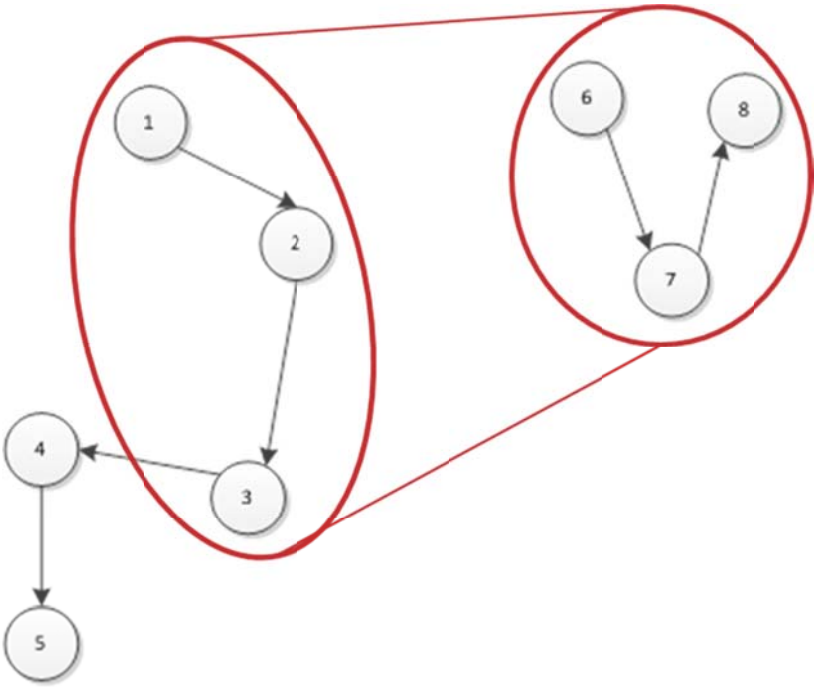



Fig. 1.6 Sub-graph comparisons

## 1.4. Thesis Contributions

This body of work provides the following contributions.

- An efficient method for coarsely registering two images utilizing the edge maps and frequency domain analysis.

- A new comprehensive image registration method that fuses feature and area-based approaches. The proposed method will involve a 3-stage approach where the validity of each stage is considered before continuing.

- An effective registration quality measure for authenticating the results of a particular registration method is proposed.

- A novel graph-based region descriptor is proposed for the identification of unique image features. The structural elements of the region descriptor, such as clustering and node arrangement, will be outlined in detail.

- Propose a nominal similarity metric and matching process for the graph-based region descriptor. The similarity score is designed to provide a large disparity between matched and non-matched region descriptors.

- Present a simple method for evaluating the different registration phases.

### 1.5. Outline

The remainder of the thesis is organized as follows. Chapter 2 provides a review of pertinent literature that is directly related to the proposed methods. Chapter 3 introduces a new graph-based region descriptor that acts as the basis for the proposed registration approach. Chapter 4 summarizes a registration technique that identifies registration control points through a 2-phase process involving normalized cross-correlation template matching and invariant feature points. Chapter 5 gives a thorough description of the proposed multi-stage registration approach. Chapter 6 provides results and discussion while Chapter 7 delivers the final conclusions and future work.

# 2. LITERATURE REVIEW

## 2.1. Image Registration

Image registration is an important processing task that serves many purposes. In medical image analysis the registration from multiple types of images provide an effective means for diagnosis [14,15] while applications involving remote sensing can benefit from the fusion of registered information [16]. Similarly, computer vision tasks such as image stitching [17] utilize automatic registration to accomplish their goal.

The process of image registration attempts to determine the homography between two images. By determining the geometric relationship between data sets, one image can be projected into the same perspective as another using the estimated transformation parameters. Image alignment can help determine mutual content between images or, conversely, identify differences which may be useful for a particular application, such as disaster assessment.

Registration algorithms can be broadly classified as feature-based, area-based, or Fourier-based approaches [18]. In the feature-based methods, distinct image features such as corners [19], gradient edges [20], or shape descriptors [21] are used to define the geometrical mapping between images. Feature-based approaches rely on the detected feature's invariance to affine, rotational, and translational transformations. With area-based methods, pixel intensities are compared directly for a sub-region of an image. In such a scenario, a similarity measure is coupled with an optimization algorithm in an attempt to identify the closest mapping of pixels [22,23]. Fourier-based registration relies

on the signal properties in the frequency domain to recover image characteristics such as translation, rotation and scale.

### 2.1.1. Feature-based Registration

With feature-based registration approaches, image attributes such as the windows of a building [24], edge-map properties of satellite images [25], or the shape of an object [26] are utilized to construct the feature space and detect potential control points. Examples of feature-based approaches include the method proposed by Guo [27] which utilizes SURF interest points and Lin [28] where Harris corner points are identified as control points. Both approaches have provided excellent results but remain computationally challenging. With the selection of control points, an algorithm is required to search the space of valid transforms. Current methods rely on some modified method of Particle Swarm theory, such as in [29-30].

#### 2.1.1.1. Control Points from SURF and Harris Corner Interest Points

In [27], Guo *et al* propose an algorithm for using SURF-detected interest points along with the Piecewise Linear Transformation (PLT) to create a network of connected triangles. The triangles of networks across images are compared and image registration parameters are calculated per matching triangle. SURF feature points are detected while corresponding descriptors are calculated and a dominant orientation for each point is derived. The Nearest and Second-nearest Neighbor Iterative (NSNI) algorithm is applied to determine the two feature points in the query image which are the top two best matches for a feature point in the reference image. Let $d_{ij}^1$ represent the descriptor distance for the top match descriptor of feature point $i$ in the reference image and feature point $j$ of the query image, and $d_{ik}^1$ be the distance to the second best matching feature point $k$ of the query image. The matched pairs are rejected if they satisfy $\frac{d_{ij}^1}{d_{ik}^1} > 0.65$. Using the SURF points as vertices, triangulation is accomplished using the Delaunay Triangulation

algorithm. Each triangle of the reference image contains the points $(x_i, y_i)$ that match to the query image at points $(X_i, Y_i)$, where $i = 1,2,3$. The relationship for the x-coordinates can be described with a simple affine transformation. The updated coordinates for the query image are computed using bilinear polynomial interpolation from the two transformation equations. An example triangle map is given in Fig. 2.1.

Fig. 2.1 Registration points identified through SURF keypoints and triangulation [27]

The Harris corner detector is a common technique for identifying control points. In [28], the authors present a multi-scale Harris detector that is used in conjunction with the Canny edge-map to isolate control points. Initially, Harris corner feature points are recognized by calculating the first order gradients of each pixel, constructing a square matrix then analyzing the eigenvalues of the matrix. For a given scale, $\sigma_i$, the matrix is created as described in the following paragraph.

For the response for all scales, local maxima and non-maximal suppression are used in scale-space to determine the candidate points that represent corners. Canny edge detection is then used to recognize edges within the image and control points are determined by calculating the edge correlation between the detected corner point and the closest edge from the edge-map. These control points are matched across the images

where $(x, y)$ is the control point in the reference image and $(x', y')$ is the matched point in the query image. The transformation between points is given as,

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = s \begin{bmatrix} \cos\theta & -\sin\theta & 1 \\ \sin\theta & \cos\theta & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}. \tag{2.1}$$

Using the two sets of control points across both images, the four unknown quantities can be derived from the following equations:

$$s = \frac{x_2' - x_1'}{\cos(\theta)(x_2 - x_1) - \sin(\theta)(y_2 - y_1)} \tag{2.2}$$

$$\theta = \tan^{-1}\left[\frac{(x_2 - x_1)(y_2' - y_1') - (x_2' - x_1')(y_2 - y_1)}{(y_2 - y_1)(y_2' - y_1') - (x_2' - x_1')(x_2 - x_1)}\right] \tag{2.3}$$

$$\Delta x = x_1' - s(x_1 \cos\theta - y_1 \sin\theta) \tag{2.4}$$

$$\Delta y = y_1' - s(x_1 \sin\theta - y_1 \cos\theta) \tag{2.5}$$

Feature-based registration is limited by the properties of the given feature point. In the strict sense, the Harris corner detector is not scale invariant; however additional research has proposed a multi-scale detector for Harris corners [31]. Furthermore, the Harris corner feature point is not discriminative enough to provide a strong one-to-one mapping between matched features of two images. SURF feature points are limited by the utilized matching technique. Several methods are proposed, but the most common approaches use a nearest neighbor approach [32]. The disadvantage of the nearest neighbor matching is the reliance upon thresholds. The matching thresholds will vary across applications. Similar to the Harris corner keypoints, the SURF feature points may not be discriminative enough in certain situations, such as images with many repeated patterns.

### 2.1.1.2. Control Point Search Algorithms

Often the control-point search space is too large to justify an exhaustive search therefore an optimization algorithm is a common approach for the selection of points. In recent literature, the most common algorithm is some variation of the Particle Swarm Theory by Kennedy and Eberhart [33]. In [29], Wang *et al* provides a registration technique for automatic image registration using SURF interest points and the Adapted Particle Swarm Optimization method. SURF feature points are calculated for 4 octaves in scale-space and provide the initial set of control points. In the standard Particle Swarm Optimization (PSO) algorithm, a random set of points is chosen as potential solutions. If the search space is $D$ dimensional, each particle $i$ has a velocity, $v_i \in \mathbb{R}^D$ and position $x_i \in \mathbb{R}^D$vectors. For each iteration, the position and velocity vectors are updated according to,

$$x_{id}^{t+1} = x_{id}^{t} + V_{id}^{t+1} \tag{2.6}$$

$$V_{id}^{t+1} = \omega V_{id}^{t} + c_1 r_1 (p_{id} - x_{id}^{t}) + c_2 r_2 \left(p_{gd} - x_{id}^{t}\right) \tag{2.7}$$

where $c_1$ and $c_2$ are acceleration constants, $p_{id}$ is the best solution for particle $i$, and $p_{gd}$ is the best solution for all particles. As an improvement to the original PSO algorithm, the Elitist Learning Strategy (ELS) is applied to compute the global best,

$$P^d = P^d + \left(x_{max}^d - x_{min}^d\right) \cdot G(\mu, \sigma^2), \tag{2.8}$$

where $G(\cdot)$ is a Gaussian random variable with mean, $\mu$, and standard deviation, $\sigma^2$. APSO is then iteratively applied along with using the Hausdorff distance as the fitness function.

The control point search algorithms pose the challenge of converging to local minima or maxima. Several variations exist for the most popular approaches where local

convergence is mitigated, but often this is at the cost of computational efficiency and ultimately total execution time. The most complicated approaches may not be suitable for real-time applications.

### 2.1.2. Area-based Registration

In contrast to feature-based approaches, intensity-based image registration directly exploits the pixel intensities and their distributions. The same optimization techniques may still apply, such as the particle swarm theory; however, a new set of similarity measures must be employed. The functions attempt to relate the statistical models of the pixels.

#### 2.1.2.1. Entropy-based Approach

Recent work in [34,35] provides an example of the statistical analysis required for medical image intensity-based registration. Given two images, $R$ and $Q$ with a partial overlap. Let $u$ represent the set of pixels in $R$ of the overlapping region, and $v$ as the set of pixels in $Q$. The entropy conditional probabilities are then defined as,

$$ECP(v|u) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(u_i, v_j) \log p(v_j|u_i) \tag{2.9}$$

$$ECP(u|v) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(u_i, v_j) \log p(u_i|v_j) \tag{2.10}$$

The modified conditional entropy is then expressed as,

$$MCE = ECP(u|v) + ECP(v|u) \tag{2.11}$$

Using the MCE as an objective function for the optimization problem, the registration is completed on the statistical distribution of the pixel intensities.

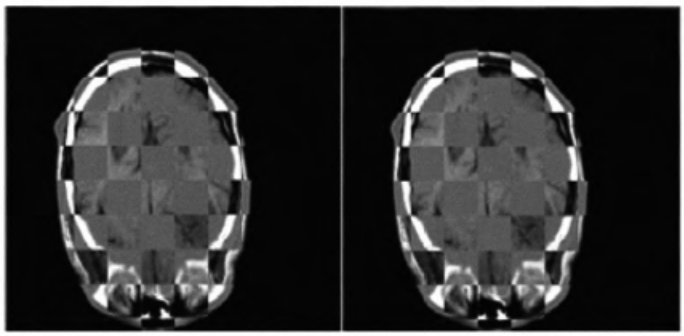


Fig. 2.2 Registration result using MCE [34]

Entropy-based methods are susceptible to variations in pixel intensity. Furthermore, the probabilistic pixel-wise comparisons are not rotational and scale invariant, which may work well for medical images, however the area-based approaches will not be appropriate for scenes before and after a disaster, captured from different perspectives and lighting conditions.

#### 2.1.2.2. Image Gradient Analysis

Shams *et al* [36] formulated a gradient-intensity measure for medical images. The gradient intensity is defined according to,

$$\Gamma(\phi,\theta) = \int\int\int h_\phi(x) h_\theta(x) f(x) dxdydz \tag{2.12}$$

$$h_\phi = \begin{cases} 1, \phi - \Delta\phi \le \phi_g < \phi + \Delta\phi \\ 0, \quad else \end{cases} \tag{2.13}$$

$$h_\theta = \begin{cases} 1, \theta - \Delta\theta \leq \theta_g < \theta + \Delta\theta \\ 0, \qquad\qquad\qquad else \end{cases} \tag{2.14}$$

While the normalized gradient intensity is expressed as,

$$\Gamma'(\phi, \theta) = \frac{\Gamma(\phi, \theta)}{max[\Gamma(\phi, \theta)]} \tag{2.15}$$

Two images are registered by computing the mutual information between normalized gradient intensity matrices. The resulting metric defines the directional similarity between the images. The mutual information comparison is invariant to translation and scale, while it is maximized when the two images are aligned rotationally. Rotational parameters are solved using the following optimization problem,

$$[\alpha_t, \beta_t, \gamma_t] = \min_{\alpha,\beta,\gamma} MI(G_h^F, G_h^{M(T)}). \tag{2.16}$$

Here, $\alpha_t, \beta_t, \gamma_t$ are the rotation parameters, $MI(\cdot)$ is the mutual information function and $G_h^F, G_h^{M(T)}$ are the gradient intensity matrices for the reference image $F$ and transformed query image, $M(T)$. The translational parameters are estimated in a similar optimization problem using image intensities for a given voxel. The proposed image gradient method couples gradient analysis with an entropy measure; however the approach will be greatly affected by noise and other variations in pixel intensity.

### 2.1.3. Image Similarity Measures

The ability for a registration problem to converge is dependent on the objective function used to evaluate the quality of the transformation. Much work has been done to associate similarity measures to registration problems.

### 2.1.3.1. Color Similarities

In [37], a measure based on color and structural information is proposed. First a structural similarity (SSIM) function is defined as,

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{2.17}$$

where $C_1$ and $C_2$ are constants, $\mu_x$ and $\mu_y$ are the means of some defined window in each image, $\sigma_x$ and $\sigma_y$ are the variances. The result from the $SSIM$ will be in the interval, $[-1,1]$, where 1 is the ideal case where the two windows are equal; $w_x = w_y$. Given a pixel $P = \begin{bmatrix} R \\ G \\ B \end{bmatrix}$ in the RGB color-space, conversion to the HSI (hue/saturation/intensity) color-space is accomplished using the following transformations.

$$H = \begin{cases} \cos^{-1}\left(\dfrac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}}\right), B \le G \\ 2\pi - \cos^{-1}\left(\dfrac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}}\right), B > G \end{cases} \tag{2.18}$$

$$S = 1 - \left(\frac{3}{R+G+B}\right) min(R,G,B) \tag{2.19}$$

$$I = \frac{R+G+B}{3} \tag{2.20}$$

If a pixel is scaled by a constant $k$, $P' = \begin{bmatrix} kR \\ kG \\ kB \end{bmatrix}$, the Hue and Saturation properties remain unchanged and the intensity is scaled by $k$; $H' = H$, $S' = S$, and $I' = kI$. The correlation coefficient for two RGB color vectors, $f$ and $g$, is defined as,

$$r(f,g) = \frac{\langle f,g \rangle}{\|f\|\|g\|} \tag{2.21}$$

The intensity similarity is then defined as

$$IS(f,g) = 1 - \frac{|f_r - f_g - f_b - g_r - g_g - g_b|}{765} \tag{2.22}$$

The color similarity is derived from both the correlation coefficients, intensity similarity function, and two predefined weights, $\alpha_1$ and $\alpha_2$.

$$CS(f,g) = \alpha_1 r(f,g) + \alpha_2 IS(f,g), \alpha_1 + \alpha_2 = 1 \tag{2.23}$$

Lastly, the color fused function is defined as,

$$CSM(A,B|F) = \lambda_1 CS(A,F) + \lambda_2 CS(B,F) \tag{2.24}$$

$$\lambda_1 = \frac{SSIM(A,F)}{SSIM(A,F) + SSIM(B,F)} \tag{2.25}$$

$$\lambda_2 = 1 - \lambda_1 \tag{2.26}$$

where $A$ and $B$ are the two images and $F$ is the fused image of $A$ and $B$.

#### 2.1.3.2. Entropy Similarities

Lei *et al*, have presented a similarity measure for web images with the intent of computational efficiency and high accuracy [38]. The similarity between two images, $X$

and $Y$, can be determined by using the HSI histograms as the probability distribution functions for each image. The Shannon entropy for image $A$ is defined as,

$$H(A) = \sum_i p_i \log_2 \frac{1}{p_i} \tag{2.27}$$

where $p_i = P(a = i)$, is the probability that pixel $i$ is the grayscale value of $a$. The proposed similarity measure based on the Shannon entropy is then expressed as,

$$I(A, B) = H(A) + H(B) - H(A, B). \tag{2.28}$$

In 2.28, $A$ and $B$ are the images and $H(A, B)$ is the entropy of the joint probability, defined as,

$$H(A, B) = \sum_{a,b} p_{AB}(a, b) \log_2 \frac{1}{p_{AB}(a, b)} \tag{2.29}$$

$I(A, B)$ will be within the interval $[-1,1]$ with a value of 1 for no dissimilarity between images $A$ and $B$.

### 2.1.4. Fourier Transform Registration

Frequency domain analysis is commonly used for determining registration parameters due to the Fourier Shift Property [39,40]. Given grayscale images, $f(x, y)$ and $g(u, v)$, where the two images only differ by a translation, $u = x + \Delta x$ and $v = y + \Delta y$, the Fourier registration approach attempts to solve for $\Delta x$ and $\Delta y$ through the use of the cross-power spectrum which is defined as,

$$\mathbb{C} = \frac{\mathcal{F}\{f\} \cdot \mathcal{F}\{g\}^*}{|\mathcal{F}\{f\}\mathcal{F}\{g\}^*|}, \tag{2.30}$$

where $\mathcal{F}\{f\}$ is the Fourier transform of image $f$ and $\mathcal{F}\{g\}^*$ is the Fourier complex conjugate of image $g$, while the $\cdot$ operator is the element-wise multiplication of the image pixels. The translation parameters are recovered by searching the normalized cross-correlation matrix that is computed from the inverse Fourier transform of the cross-power spectrum, $c = \mathcal{F}^{-1}\{\mathbb{C}\}$. The peak of the cross-power spectrum is given as,

$$(\Delta x, \Delta y) = \max_{x,y}(c). \tag{2.31}$$

Rotational and scale differences between images can be recovered using the same cross-power spectrum principals, however the grayscale images must first be converted using the log-polar transformation. The log-polar transformation that maps Cartesian coordinates to log-polar coordinates, illustrated in Fig. 2.3, is defined as,

$$(\theta, r) = \left(\tan^{-1}\frac{y - y_c}{x - x_c}, \log\sqrt{(x - x_c)^2 + (y - y_c)^2}\right) \tag{2.32}$$

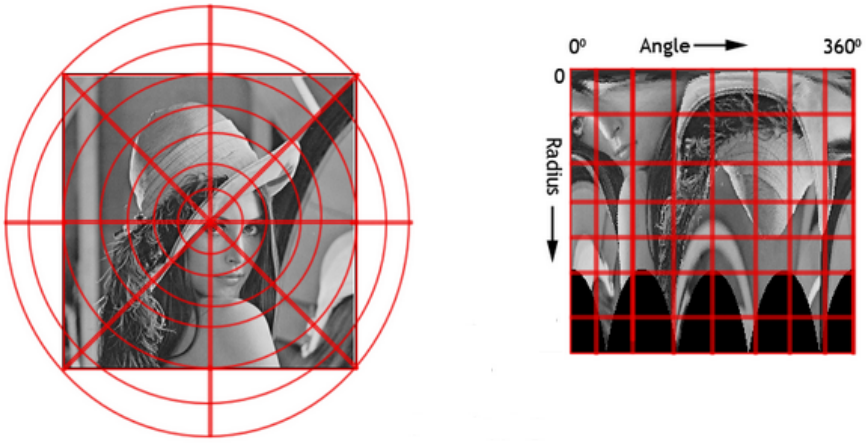



Fig. 2.3 Polar transformation

The rotation and scaling parameters, $\Delta r$ and $\Delta\theta$, are then computed using the cross-power spectrum of the log-polar images. Let $c_{lp}$ be the resulting normalized cross-correlation matrix using the log-polar grayscale images. Then the registration parameters are solved using,

$$(\Delta r, \Delta\theta) = \max_{r,\theta}(c_{lp}) \tag{2.33}$$

Log-polar FFT registration is a popular approach [40-42], however the accuracy is affected by noise and other photometric variations such as object occlusion. To mitigate these deficiencies, the method is often used as a coarse registration pre-processing step where more modern techniques, such as feature-based registration, is used to achieve sub-pixel registration [43]. Moreover, the Fourier-based approaches are not accurate for large translations and rotations.

## 2.2. Image Features

At the core of reliable image matching is the accurate identification of features within an image. The detected features should be invariant to geometric and photometric effects. Also, the detected features should be unique enough so that ideally there will exist a strict one-to-one mapping across images. In the following section, popular invariant feature descriptors are reviewed, followed by a section about region descriptors.

### 2.2.1. Feature Point Descriptors

Many feature point descriptors have been proposed, however, the most common include SIFT, SURF, Harris corner, and invariant moments. Each has its strengths and weaknesses which are summarized in the following sections.

#### 2.2.1.1. SIFT Feature Points

The Scale Invariant Feature Transform (SIFT) feature point descriptor proposed by Lowe [44] has long been the reference metric for comparing other point descriptors. The scale-space is created by using a pyramid of Gaussian-blurred images. Let $L(x, y, \sigma)$ represent the filtered image for a given scale, $\sigma$.

$$L(x, y, \sigma) = I(x, y) * G(x, y, \sigma) \tag{2.34}$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2.35}$$

where $G(x, y, \sigma)$ is the Gaussian function that is convolved with the original image, $I(x, y)$. The scale-space is constructed by iteratively increasing the size of the Gaussian kernels. The difference of adjacent filtered images is computed to create the Difference-of-Gaussian (DoG) pyramid. Within this space, candidate feature points are chosen from a 3x3x3 neighborhood by searching the DoG pyramid. For each identified keypoint, the gradient magnitude and orientation histograms are computed from the following,

$$m(x, y) = \sqrt{[L(x+1, x) - L(x-1, y)]^2 + [L(x, y+1) - L(x, y-1)]^2} \tag{2.36}$$

$$\theta(x.y) = \tan^{-1} \frac{L(x, y+1) - L(x, y-1)}{L(x+1, x) - L(x-1, y)} \tag{2.37}$$

The resulting descriptor is a 128-dimensional vector composed of both histograms. Several SIFT optimizations have been proposed to overcome some of the shortcomings of the original algorithm. In [45], the descriptor dimensionality is reduced by employing Principal Component Analysis (PCA) on the image patch about each feature point, resulting in a 36-dimensional vector.

### 2.2.1.2. SURF Feature Points

In [46], Bay proposed a novel alternative to SIFT called Speedup Robust Features (SURF) which aimed to compute multi-scale feature points which are invariant to scale, rotation and translational deformations. The algorithm relies on integral images for fast computation of discrete image integrals, which is defined in (2.38).

$$I_{\Sigma}(x,y) = \sum_{i=0}^{x}\sum_{j=0}^{y} I(x,y) \tag{2.38}$$

By utilizing the integral image, the area within a bounded region (A,B,C,D) of the original image can be computed using four memory accesses and three additions.

$$\Sigma = I_{\Sigma}(A) - I_{\Sigma}(B) - I_{\Sigma}(C) + I_{\Sigma}(D) \tag{2.39}$$

Interest points are determined by calculating the determinant of the Hessian matrix for a particular pixel location (x,y).

$$\begin{aligned} &\det\left(\mathrm{H} = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix}\right) \\ &= L_{xx}(x,\sigma)L_{yy}(x,\sigma) - 0.9\left(L_{xy}(x,\sigma)\right)^2 \end{aligned} \tag{2.40}$$

Where $L_{xx}(x,\sigma) = \frac{\partial^2}{\partial x^2} G(\sigma) * I(x,y)$ for a given scale, $\sigma$. The Gaussians are estimated using discrete box filters. The following images in Fig. 2.4 illustrate the discrete box filters for $L_{yy}$ and $L_{xy}$.

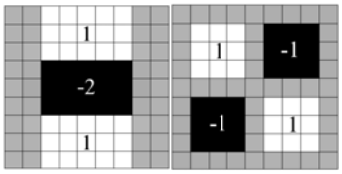



Fig. 2.4 Box filters for $L_{yy}$ and $L_{xy}$

To create the scale-space representation, the discrete box filters are increased in size while the image maintains a constant size, which is in contrast to the typical Gaussian pyramid scheme where the Gaussian kernel is kept constant and the image size is scaled. With SURF, the computational time for any filter size remains constant, thus increasing the efficiency over SIFT. From the scale-space, a search is performed in a 3x3x3 neighborhood while non-maximal suppression identifies candidate feature points. For each detected point, a descriptor is calculated from haar wavelets and their dominant orientations.

$$d = \begin{bmatrix} \Sigma d_x \\ \Sigma d_y \\ \Sigma |d_x| \\ \Sigma |d_y| \end{bmatrix}, \quad d \in \Re^{64} \tag{2.41}$$

In (2.41), $\Sigma d_x$represents the haar response in the x-direction and $\Sigma d_y$is the response in the y-direction about a 4x4 window centered at the detected keypoint.

### 2.2.1.3. Invariant Moments

Invariant moments are yet another popular feature often used to describe image content [47-49]. Similar to the SIFT and SURF descriptors, the moment descriptor is invariant to rotation, scale and translation deformations. The general form for the descriptor is expressed as,

$$m_{pq} = \sum_y \sum_x x^p y^p I(x, y) \tag{2.42}$$

The most frequently used moments are the Zernike Moments where the moments defined by (2.42) are altered such that $x^p y^p$ is replaced with Zernike basis functions, resulting in the following moments,

$$Z_{nm} = \frac{n+1}{\pi} \sum_{\theta} \sum_{\rho} V_{nm} I(\rho, \theta) \tag{2.43}$$

$$V_{nm} = e^{jm\theta} \sum_{s=0}^{n-|m|/2} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \rho^{n-2s} \tag{2.44}$$

#### 2.2.1.4. Binary Point Descriptors

In recent work, binary descriptors are used as an improvement over common approaches such as SURF and SIFT. They often involve rotational, scale and translational invariance but use a binary string to represent the resulting descriptor. Traditional descriptors are represented by a series of floating point numbers, therefore significant performance increases will be realized when utilizing binary strings. Storage requirements decrease, descriptors are created with lower latencies, and matching is performed at an efficient rate due to the use of the Hamming distance metric instead of the customary Euclidean distance.

Binary Robust Invariant Scalable Keypoints (BRISK) is a binary point descriptor proposed by Leutenegger *et al* [50]. The proposed technique proposes the use of the Adaptive and Generic Accelerated Segment Test (AGAST) feature point detector which identifies corners utilizing a binary decision tree. AGAST is a derivative of the FAST corner detector. Similar to the SURF detector, AGAST exploits a scale-space pyramid to identify scale invariant keypoints.

BRISK descriptors are constructed by comparing the brightness characteristics of neighboring patches of pixels. Sampling is conducted along concentric circles as depicted in Fig. 2.5. The sampling pattern is determined from the characteristic pattern

direction that is derived from gradient comparisons between the detected feature point and points that are at least $13.67k$ pixels apart, where $k$ is the detected scale of the feature point. Once the characteristic direction has been determined, points are sampled that are within $9.75k$ pixels of the key point. Bit $b$ of the descriptor is defined as,

$$b = \begin{cases} 1, & I(p_j^\alpha, \sigma_j) > I(p_i^\alpha, \sigma_i) \\ 0, & otherwise \end{cases} \tag{2.45}$$

where $I(p_j^\alpha, \sigma_j)$ is the Gaussian smoothed image patch for point $j$, and $I(p_i^\alpha, \sigma_i)$ is the image patch about the feature point.

Descriptor matching is accomplished using the efficient Hamming distance that is defined as number of set bits from the bit-string result of a bit-wise XOR operation on two descriptors. Given two descriptor bit-strings, $d_i$and $d_j$, the XOR result,$r = d_i \otimes d_j$, is used to define the following Hamming distance.

$$H = \sum_r r_i \tag{2.46}$$

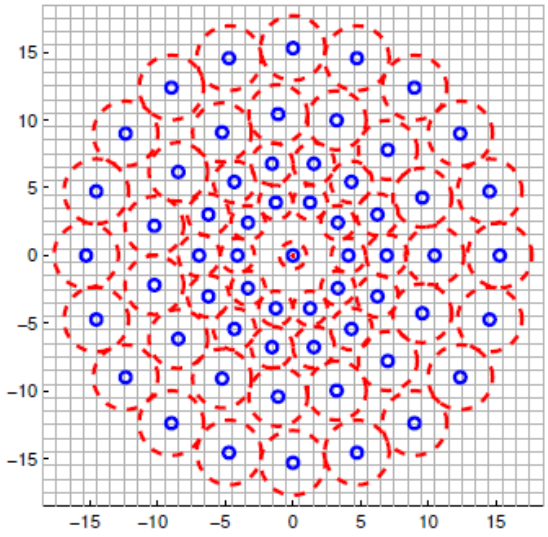



Fig. 2.5 BRISK sampling pattern for $k$ = 1 [50]

The authors in [50] show that BRISK offers accuracy that is comparable to SURF, however BRISK is shown in some circumstances to be an order of magnitude faster than its counterpart.

### 2.3. Region Descriptors

In certain situations, local features may not be unique enough to offer a one-to-one mapping to another image. Such scenarios include building detection and matching. With large city buildings, many patterns are repeated in terms of texture and spatial relationships. Windows are often not distinctive enough for matching individual features. Another example would be the comparison of two images before and after a disaster. Post-disaster, the objects or scene may have been greatly altered and therefore the gradient information or haar wavelet responses of local feature points may not be suitable for comparison. To overcome these issues, larger areas including several image features are evaluated to develop region descriptors. Region descriptors may offer more detail and information about an image area than local feature descriptors.

Optimization algorithms are a common tool for grouping local feature points into a single region descriptor. In [51], Wang *et al* propose a new binary PSO algorithm for feature point clustering to identify regions of interest, while Awais and Mikolajczyk have proposed a region descriptor by grouping the feature points along the dominant lines of an image [52].

Dominant lines are determined by first using an edge detector or by applying a segmentation method and extracting the cluster boundaries. The authors chose to use the watershed segmentation along with RANSAC to estimate the line parameters which best approximate the object boundaries. A total of 4 parameters are estimated; line mid-point, $(x_c, y_c)$, length of the line, $l$, the distance of the perpendicular line to the origin of the image, $\rho$, and the orientation of the perpendicular line, $\theta$. The feature points about a line are determined for a circular region centered at $(x_c, y_c)$ with a diameter of $l$. Similar to SIFT, the area is encoded as a histogram of gradient orientations. As an additional

feature, the authors propose the usage of the vertices connecting two dominant lines. Like the previous feature, an area around each of the vertices is encoded as a histogram of gradient orientations. The coupling of two endpoint features and midpoint are used as a single region feature for that boundary.

In a similar approach, Liu *et al* have presented a method for extending the SIFT or GLOH feature point descriptors to line and curve descriptors [53]. First, feature points are detected using the standard SIFT algorithm. For the 16x16 region centered at each descriptor, the average gradient orientation is calculated. If the average gradient orientation is the same across multiple feature points, the descriptors lay on the same line, otherwise the point represent either a curve or closed region. To compute a descriptor for a line, curve or closed region that is composed of $N$ feature points, first compute the average descriptor vector.

$$M = \frac{1}{N}\sum_{i=1}^{N} V_i = [M^1, M^2, .., M^N] \in \mathbb{R}^n, V_i \in \mathbb{R}^n \tag{2.47}$$

Next the standard deviation vector is calculated.

$$S^j = \frac{1}{N}\sum_{i=1}^{N} \left(V_i^{\,j} - M^j\right)^2, j = 1, .., n \tag{2.48}$$

A descriptor vector is created using the normalized coupling of the mean and standard deviation vectors. This new vector is called the mean and standard deviation descriptor (MSD).

$$MSD = \left[\frac{M}{\|M\|} \quad \frac{S}{\|S\|}\right] \in \mathbb{R}^{2n} \tag{2.49}$$

As an alternative to the MSD, a moment descriptor (MD) is computed using the following four moments from the average gradient orientations.

$$M_1^j = \frac{1}{N}\sum_{i=1}^{N}\left(V_i^j - M^j\right), j = 1,..,n \tag{2.50}$$

$$M_2^j = \frac{1}{N}\sum_{i=1}^{N}\left(V_i^j - M^j\right)^2, j = 1,..,n \tag{2.51}$$

$$M_3^j = \frac{1}{N}\sum_{i=1}^{N}\left(V_i^j - M^j\right)^3, j = 1,..,n \tag{2.52}$$

$$M_4^j = \frac{1}{N}\sum_{i=1}^{N}\left(V_i^j - M^j\right)^4, j = 1,..,n \tag{2.53}$$

The MD is then created as the normalized vector of the four moments.

$$MD = \left[\frac{M_1^T}{\|M_1\|} \quad \frac{M_2^T}{\|M_2\|} \quad \frac{M_3^T}{\|M_3\|} \quad \frac{M_4^T}{\|M_4\|}\right] \in \mathbb{R}^{4n} \tag{2.54}$$

In addition to the MSD and MD descriptors, a Fourier descriptor (FD) is proposed using the first $n$ Fourier coefficients.

$$F_i = [F_i^1, F_i^2, .., F_i^n] \in \mathbb{R}^n, i = 1,2,34 \tag{2.55}$$

$$F_1^j = \mathcal{F}_1\{V_1^j, V_2^j, .., V_n^j\}, j = 1,2,..,n \tag{2.56}$$

$$F_2^j = \mathcal{F}_2\{V_1^j, V_2^j, .., V_n^j\}, j = 1,2,..,n \tag{2.57}$$

$$F_3^j = \mathcal{F}_3\{V_1^j, V_2^j, .., V_n^j\}, j = 1,2,..,n \tag{2.58}$$

$$F_4^j = \mathcal{F}_4\{V_1^j, V_2^j, .., V_n^j\}, j = 1,2,..,n \tag{2-59}$$

Finally, the FD is defined as the normalized vector of Fourier coefficients.

$$FD = \left[\frac{F_1^T}{\|F_1\|} \quad \frac{F_2^T}{\|F_2\|} \quad \frac{F_3^T}{\|F_3\|} \quad \frac{F_4^T}{\|F_4\|}\right] \tag{2.60}$$

## 2.4. Summary

This chapter has provided a broad overview of various image registration techniques that can be categorized into feature, area, or Fourier-based. Feature-based approaches will attempt to use matching image features to estimate an appropriate homography, whereas area-based methods compare the image pixels directly along with evaluating an objective function for converging onto a solution. Fourier-based approaches take advantage the shift properties of a signal in the frequency domain. Translation offsets can be computed from a Fourier-transformed image pair while rotation and scale can be determined from the log-polar transformed image pair. Lastly, several image comparison metrics are presented. The literature review also includes detailed explanations of common feature point algorithms, SIFT, SURF and BRISK. For completeness, recently proposed region descriptors are discussed.

# 3. GRAPH-BASED INVARIANT REGION DESCRIPTOR

Identification of invariant image descriptors is an integral task for many computer vision applications such as image registration, object recognition, and object tracking. The detected features should be invariant to geometric transformations such as rotation and translation, as well photometric variations due to differing lighting conditions. In this work, we propose a unique and effective region descriptor that couples invariant features and texture information. The descriptor relies on spatial relationships of invariant SURF features to create a graph-based descriptor for image matching. Additionally, a novel method is proposed for matching region descriptors through the definition of an efficient similarity measure that couples invariant features and their spatial relationships. Several examples are presented to illustrate the effectiveness of the proposed region descriptor while the results of the proposed approach outperform SURF feature point matching.

Typical approaches that rely on invariant features can be thought of in three phases. Initially, invariant features are detected, followed by invariant feature description, and then feature matching is performed between two images. Common keypoint detection algorithms include SIFT [44], which uses a scale-space from a pyramid of difference of Gaussian images, and the Harris Corner Detector [54] that attempts to identify image corners through the analysis of eigenvalues for the first order gradients of local image neighborhoods. The original Harris Corner Detector algorithm is not scale invariant and thus several variations have been proposed [55,56], however these proposed methods do not address the issue where one invariant feature of the reference image maps to several features in the query image.

Due to the scale, rotation, and translation invariance of the Scale Invariant Feature Transform (SIFT) descriptor, much work has been done based on the SIFT detector and feature point descriptor, however the complexity of the algorithm limits its applications. In [46], Bay proposes an alternative to SIFT called Speeded-Up Robust Features (SURF) that is also scale, rotation and translation invariant. SURF is advantageous over SIFT in several areas. The algorithm utilizes integral images and discrete box filters for fast, integer-based computation of the scale-space. Moreover, the SURF descriptor is based on an integer approximation of the Haar wavelet responses within a local neighborhood of each feature point. Due to these enhancements and promising results, SURF has quickly become the basis for many computer vision applications [57-60]. Although SURF greatly reduces the complexity of SIFT, the discriminative power of the SURF descriptor does not always ensure a one-to-one mapping between invariant feature points from two images. One such scenario where this issue may arise is that of buildings whose structure includes many repetitive characteristics, such as windows. It is the aim of this work to use SURF detected feature points to create a more discriminative descriptor that will have a one-to-one mapping across images. In doing so, the overall solution is computationally efficient and may be a suitable candidate for real-time applications or easily mapped to hardware using reconfigurable logic.

In this work, a novel region descriptor is proposed that is invariant to both photometric and geometric variations. The computational efficiency of the SURF algorithm is leveraged to create a region descriptor that couples spatial, textural, and color information. In doing so, it is shown that the discriminative power of the proposed descriptor is superior to the SURF feature points alone. The overall structure of the region descriptor forms a connected graph which provides an efficient structure for rapid matching. Additionally, a unique similarity measure is proposed for comparing two region descriptors.

### 3.1. Graph-based Descriptors

Graph-based analysis has recently become a popular method for determining global features from invariant local characteristics. The overall discriminative power of invariant feature points is potentially increased if the graph is also invariant to geometric or photometric variations. Simacek and Unsalan proposed a connected graph from SIFT points in [61] to detect buildings in an urban scene, while Kisku et al used a similar approach for a facial recognition system [62]. In both instances the authors show an increased matching accuracy when comparing subgraphs from two images than using the feature points alone. Neither approach utilizes color information while both methods rely on the complex SIFT descriptor which may limit the scope in which the algorithms may be used.

### 3.2. Proposed Graph-based Descriptor

Image matching requires the identification of common features or regions between images. If enough similarity exists, the scene or objects are matched. The detected features of each image should be distinct and the method should be invariant to photometric and geometric inconsistencies. The proposed algorithm creates a descriptor from the spatial and color relationships of invariant descriptors. Because of object deformations and occlusions, the matching process must calculate similarities between regions to determine all commonalities between images.

To mitigate these issues, it is the aim of this paper to propose a region descriptor that is formed from a connected graph that contains SURF feature points as its vertices. Fig. 3.1 provides an overview of the proposed method for feature detection and matching.

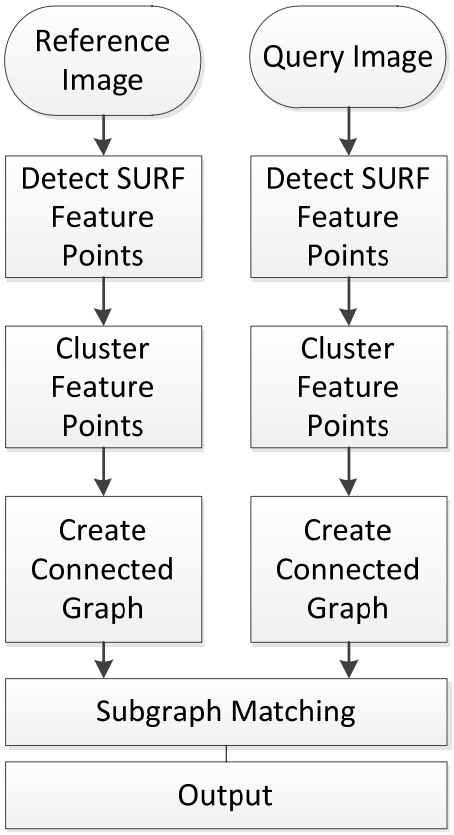



Fig. 3.1 Region descriptor overview

For a given reference and query image, the SURF feature points are computed then a clustering process is used to identify similar feature points. The clustering algorithm ensures all feature points within a group share common characteristics and they are invariant to rotation, translation and affine transformations. Each cluster will also share common color properties as determined by the area's color distribution. A connected graph is then constructed from each feature point within a cluster. To accomplish this, a proposed similarity measure is used to exploit the spatial and textural characteristics of each feature point while Dijkstra's algorithm is employed to create a connected graph of feature points.

Matching two images involves matching region descriptors across images. Since each descriptor is a connected graph, sub-graph matching schemes are explored. Often exact matches are not possible due to deformations, occlusions or variations in color; therefore the problem of region matching must be accomplished using a proposed similarity measure. This measure compares two graphs on the basis of the individual nodes and their spatial relationships in feature-space.

### 3.2.1. Clustering

Initially feature points are detected using SURF. These provide the basis for the subsequent creation of directed graphs that will represent the proposed region descriptor. Two methods are proposed and compared. First, color characteristics are proposed then a method for clustering based on spatial density using a k-means approach is outlined. After clustering, each group is analyzed and clusters with less than four nodes are discarded. For effective matching, at least four nodes are required, as discussed in section 3.3.

#### 3.2.1.1. Color and Descriptor Characteristic Clustering

The first method performs two passes through all feature points, comparing each color channel in the RGB color space, as well as their respective descriptors. The first pass will determine initial groupings while the second pass attempts to merge overlapping and similar clusters.

The initial feature point detection is accomplished using the SURF detector. Two 64-element descriptors, $d_k$ and $d_j$, are compared by calculating their Euclidean distance in feature space according to (3-1).

$$S_E(d_k, d_j) = \sqrt{\sum_{i=1}^{64} \left(d_k[i] - d_j[i]\right)^2} \tag{3.1}$$

Let $S_c$ be the similarity of two feature points in RGB color space defined as,

$$S_c = \sqrt{(I_R^1 - I_R^2)^2 + (I_G^1 - I_G^2)^2 + (I_B^1 - I_B^2)^2}, \tag{3.2}$$

where $I_c^x$ is the color channel for feature point x, with $c \in \{R, G, B\}$. Coupling color and descriptor similarity, a potential cluster of similar descriptors is expressed as,

$$S(i,j) = \{S(i,j) \mid S_E(i,j) \leq T_E, S_c(i,j) \leq T_C\}. \tag{3.3}$$

Let $\Psi^{nxn}$ be a search window centered at a feature point. The set of descriptors within the search window is the set $C$, defined as,

$$C = \{d_i \mid d_i \in \Psi^{nxn}\}. \tag{3.4}$$

The descriptors within a search window are clustered if 75% of the descriptors are similar to all other descriptors within the window. We define the set of clustered descriptors as

$$B_i = \left\{d \in C \mid \frac{|S(i,j)|}{|C|} \geq 0.75\right\}. \tag{3.5}$$

If we define the set of descriptors in cluster n as $d^n = \{d \mid d \in B_n\}$ then two clusters are merged if the number of common feature points between clusters is at least 2. The resulting set is defined in (3.5).

$$B_i^* = \left\{B_i \cup B_j \mid \left( \left|d^i \in B_j\right| \geq 2 \right)\right\}. \tag{3.6}$$

An example clustering of feature points is provided in Fig. 3.2 where SURF points are marked in green and the clusters are outlined in red.

a. b.

Fig. 3.2 Detected feature point clusters (a), and a close view (b) of the single cluster indicated by the blue outline in (a)

### 3.2.1.2. Spatial K-means

K-means clustering has long been exploited for many applications where clustering may be necessary [63,64]. In this work, k-means clustering is utilized to group feature points based on their spatial relationships. This will result in grouping together feature points into dense groups. The original method proposed by MacQueen [65] involves a three step process; initial cluster assignment, calculation of cluster center, and cluster reassignment. Steps 2 and 3 are repeated until the method converges to the point where clusters remain unaltered.

Given a set of $n$ observations, $S = \{x_d \mid 1 \leq d \leq n\}$, the aim of the algorithm is to produce $k$ clusters, $S = \{S_1 \cup ... \cup S_k\}$, where each cluster has an associated mean, $m_1 ... m_k$. Observations are then assigned to each cluster contingent upon the distance between the observation and the cluster means. Set $S_j$ at the next iteration of $m$ is defined as,

$$S_j^{i+1} = \{x_d \mid \|x_d - m_t^i\| < \|x_d - m_p^i\| \}\forall p \neq t \tag{3.7}$$

The algorithm has converged when the following condition is reached.

$$S_j^{i+1} = S_j^i \ \forall \mathrm{j} \tag{3.8}$$

For the proposed region descriptor, the cluster centers are initially selected at random. For a set of feature points, $S_{FP}$, the number of clusters is defined as,

$$k = \frac{|S_{FP}|}{5} \tag{3.9}$$

The denominator is chosen such that the average number of nodes per graph is 5. This value allows the region descriptor to be unique but not so large that it would inhibit descriptor matching with additional latency. Initial cluster centers are represented by the spatial coordinates of the randomly selected feature points, where initial assignments are determined by the following,

$$S_j^0 = \{x_d \mid \|x_d - x_t^0\| < \|x_d - x_p^0\| \}\forall \mathrm{p} \neq \mathrm{t}, \tag{3.10}$$

where $x_t^0$ is the feature point randomly selected as an initial cluster center. The Euclidean distance metric is used for evaluating the distance between the current cluster centers and

all other feature points. Given feature points $FP_i$ and $FP_j$ and spatial coordinates, $(x_i, y_i)$ and $(x_j, y_j)$, respectively, then the distance is calculated from,

$$d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \tag{3.11}$$

The objective function defined in (3.13) provides the scoring mechanism for the reassignment phase of the clustering. During the update stage of clustering, for each cluster $S_i$, the mean $\mu_i$ and standard deviation $\sigma_i$, are calculated for the $x$ and $y$ spatial coordinates of all feature points belonging to that cluster. The k-means clustering of invariant feature points is illustrated in Fig. 3.3.

$$\mathbb{G}_{2D}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\left[\frac{x-\mu_x}{2\sigma_x^2} + \frac{y-\mu_y}{2\sigma_y^2}\right]} \tag{3.12}$$

$$1 - \mathbb{G}_{2D}(x, y) \tag{3.13}$$

Fig. 3.3 Example clustered feature points using the k-means approach with a 2D Gaussian. Feature points indicated in the same color belong to the same cluster

### 3.2.2. Graph Structure

If we assume a cluster of feature points represent the vertices of a graph, we can formulate an algorithm for the creation of a region descriptor in the form of a connected graph. Two approaches are explored. First, Dijkstra's algorithm is used to determine the shortest path from one point to any other point in a connected graph given a list of weighted edges. If we define a starting node, a directed graph can be constructed as a shortest path between the origin node and all other nodes. Second, a greedy node selection algorithm is used to produce a graph based on the descriptor characteristics instead of spatial relationships. The greedy algorithm produces the graph by iteratively choosing graph nodes that will not create a cycle.

#### 3.2.2.1. Dijkstra's Algorithm

For the work in [11,66], the initial node of the graph is chosen as the node that is within the closest proximity of the image origin. This choice ensures consistency as to how the graphs are constructed, however a more robust approach is proposed. For a given set of $n$ feature points, $FP$, with $n$ descriptors, $\{d_1, \dots, d_n\} \in FP$, the average descriptor for that set is given as,

$$d_{avg} = \frac{\sum_{i=1}^{n} d_i}{|FP|}. \tag{3.14}$$

The initial node, $d_0$, is then chosen as the node within closest proximity, in terms of descriptor distance (3.1), to the average descriptor, $d_{avg}$.

$$d_0 = \min_n \left\| d_i - d_{avg} \right\|. \tag{3.15}$$

Let $Q_G$ be defined as the queue representing our final connected graph. The first node in the queue will be regarded as the origin of the graph. Initially the queue is empty; $Q_G = \{\emptyset\}$. We define $C_F$ as the set containing the cluster of feature points. Dijkstra's approach to feature point clustering is depicted in Fig. 3.4 and outlined as follows,

- Select origin node, $FP_0$ and push into the queue, $Q_G = \{FP_0\}$ and remove $FP_0$ from the set of feature points; $C_F = C_F \setminus FP_0$
- Let $FP_i$ represent a connected neighbor of $FP_0$. If $FP_i$ is already associated with a spatial distance, add $\|FP_0 - FP_i\|$, otherwise assign $\|FP_0 - FP_i\|$. Repeat for all remaining feature points in the cluster.
- Select $FP_j$ that achieves the minimum, $\min_{FP_j} \|FP_0 - FP_j\|$.
- Push $FP_j$ onto the queue; $Q_G = Q_G \cup FP_j$.
- Remove $FP_j$ from the feature point set; $C_F = C_F \setminus FP_j$.
- For the node at the tail of the queue, repeat from step 3 until $C_F = \{\emptyset\}$.

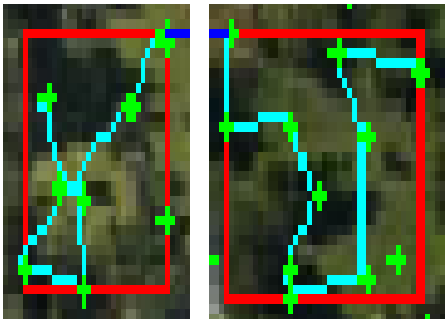


Fig. 3.4 Two example connected graphs from invariant feature points. The green markers indicate the feature points while the blue lines are the connected edges of the graph

### 3.2.2.2. Greedy Node Selection

As an alternative to Dijkstra's method, we propose the use of a greedy algorithm that is coupled by a feature point filtering scheme for reducing graph sizes. For multi-scale descriptors, such as SIFT, SURF and BRISK, the descriptor is determined by a sampling

window that is related to the detected scale. For SURF in particular, the sampling window is defined with a radius of $6s$ pixels, where $s$ is the detected scale [46]. The filtering step in the proposed selection process attempts to reduce redundant information from overlapping descriptors. Fig. 3.5 exemplifies descriptor scale overlap.

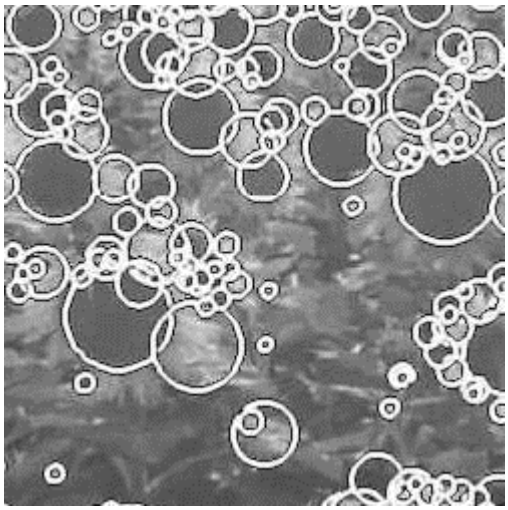


Fig. 3.5 Example SURF descriptors and their detected scales [46]

The initial feature point is selected in the same routine where the node is selected based on the descriptor distance and overall average descriptor value for that particular cluster. Given the initial feature point, $FP_0$, all feature points within the neighborhood of $6s_0$ pixels are discarded, where $s_0$ denotes the scale of the initial node. The initial node is then regarded as the current node. The descriptor distance (3.1) is computed from the current node to all other nodes. The next node chosen is the node with the smallest descriptor distance to the current node. Another iteration is started with feature point filtering. This is repeated until all feature points in a cluster are assigned to the graph or discarded. Lastly, the resulting graph is discarded if it contains less than four nodes. Example graphs are provided in the following Fig. 3.6.

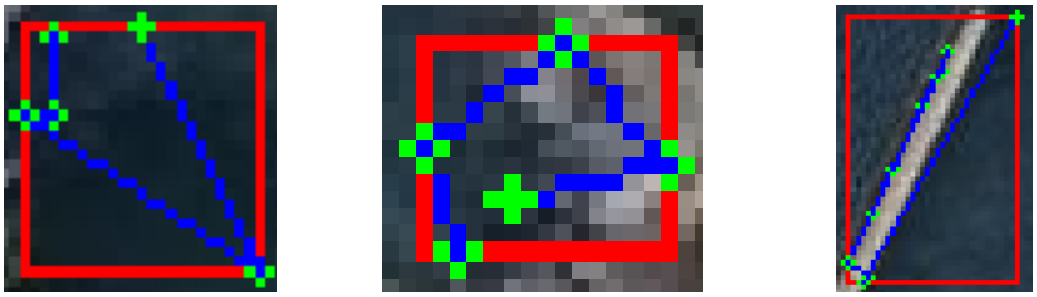


Fig. 3.6 Sample graphs created using greedy node selection with feature point filtering

### 3.3. Graph-based Descriptor Matching

For both proposed graph-based region descriptors, the method for comparing two descriptors can be regarded as finding the best matching graph. When two graphs are of differing sizes, the problem can be viewed as searching for the best sub-graph within the larger of the two graphs.

#### 3.3.1. Spatial and Descriptor Characteristic Matching

With $G_1$ and $G_2$ representing two region descriptors, region matching can be reduced to the problem of determining whether or not $G_1$ is a sub-graph of $G_2$, with $|V_2| > |V_1|, G_1 = (V_1, E_1), G_2 = (V_2, E_2)$. For simplicity, we are not searching for exact matches but rather similar sub-graphs where the similarity is a calculation based on the feature point's descriptor and spatial information of the graph nodes. The graph from $B_i^*$ is compared to the graph which spans $B_j^*$, and the lowest similarity score from the comparison is stored in a scoring matrix, $\mathcal{M}^{|B_i^*|x|B_j^*|}$, where $B_i^*$ represents a cluster of feature points.

The nodes have an associated descriptor, $d \in \mathbb{R}^{64}$and we define the similarity of two nodes, $V_i$ and $V_j$, to be $S_{FP} = \|d_i - d_j\|$. In a similar fashion, we can define the spatial

similarity between two edges, $E_i$ and $E_j$ as $S_D = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$, where $V_i = (x_i, y_i)$ and $V_j = (x_j, y_j)$. The similarity between two graphs can then be computed as,

$$S = \frac{W_1 \sum_{|G_i|} S_{FP} + W_2 \sum_{|G_i|} S_D}{W_1 + W_2} \tag{3.16}$$

where $G_i$ is the smaller of the two graphs being compared, $W_1$is the weight of the feature point descriptors while $W_2$ is the weight associated with the spatial distance between nodes. For this version, $W_1$ is set to 0.8 and $W_2$ is 0.4.

Given two graphs, $G_i$ and $G_j$ such that $|V_i| = n$, $|V_j| = m$, and $n > m$, we first compute a similarity score between $G_j$and a sub-graph of $G_i$, denoted as $G_i'$ where $G_i'$ contains the first $m$ vertices of $G_i$. In successive iterations, we extract sub-graphs of size $m$, starting from the next node in the queue of $G_i$ until we have $(n - m + 1)$ similarity scores. The smallest of these scores are chosen as the similarity score for $\mathcal{M}_{i,j}$. Once the exhaustive search is complete, $\mathcal{M}$ is traversed to find matches which are indicated by minimal scores.

If we are comparing two graphs, $G_1$ and $G_2$, where $n = |V_1|$, $m = |V_2|$, and $n \geq m$, we must make $n - m + 1$ comparison to test $G_2$ against all sub-graphs of $G_1$. In the following figure, the vertices of $G_1$are given as the set $\{V_1, .., V_5\}$ and the vertices of $G_2$ as $\{V_6, .., V_8\}$.

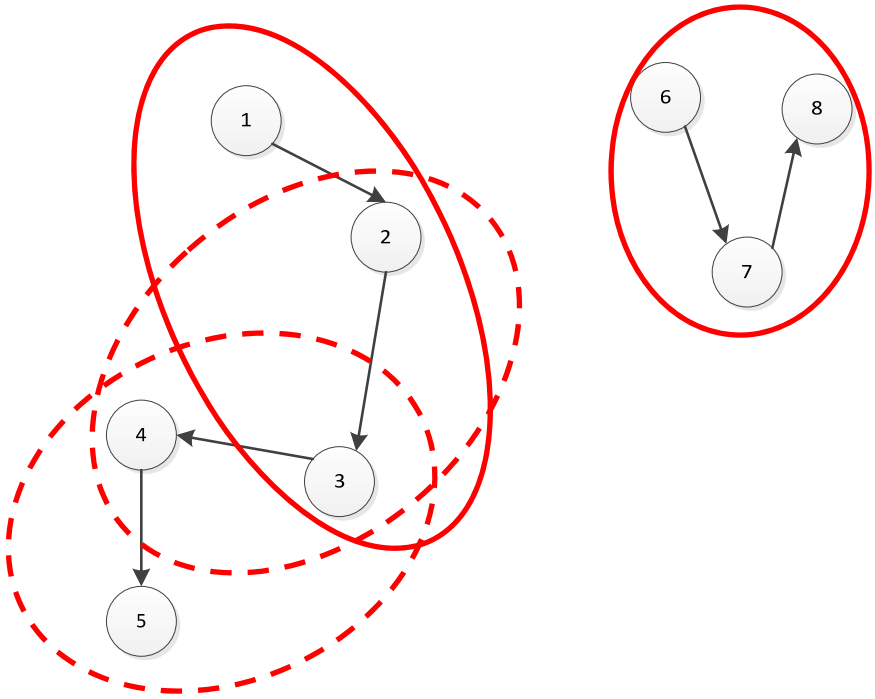



Fig. 3.7 Partial graph matching

For the illustrated example in Fig. 3.7, a matching score is generated for the comparison of sets $\{V_1, V_2, V_3\}$ and $\{V_6, V_7, V_8\}$, $\{V_2, V_3, V_4\}$ and $\{V_6, V_7, V_8\}$, and finally between $\{V_3, V_4, V_5\}$ and $\{V_6, V_7, V_8\}$. The minimal score represent the similarity between $G_1$ and $G_2$. Section 3.5 provides example results from matching the connected graph region descriptors.

### 3.3.2. Angle and Descriptor Characteristic Matching

The proposed similarity measure from equation (3.16) attempts to compare two graphs from their structural characteristics and image features; however the previously proposed similarity score is not rotationally invariant. Due to this shortcoming, it is the aim of the following similarity metric to couple the graph's angles with image features to provide a rotationally invariant score. Moreover, the metric is structured such that there are larger penalties for mismatched features and angles through a non-linear formula.

Provided a directed graph with $n$ vertices, $V = \{v_0, v_1, \dots, v_{n-1}\}$, and $n-1$ edges, $E = \{e_{01}, e_{12}, \dots, e_{(n-3)(n-2)}\}$, where $e_{ij}$ denotes the edge between vertices $v_i$ and $v_j$, we define the $n-2$ angles as $A = \{a_{01}, a_{12}, \dots, a_{(n-4)(n-3)}\}$, where $a_{ij}$ represents the angle between edges $e_i$ and $e_j$. As a result of the graph's directed edges, the angle order is critical. For a given graph, two angle descriptors can be constructed to represent the structure of the graph. First, a descriptor for a graph with $n$ nodes will be structured with $n-2$ elements where $n \geq 4$. Let $d_{\theta 1}$ be a graph descriptor composed of the ordered angles,

$$d_{\theta 1} = \begin{bmatrix} \theta_{0,1} \\ \theta_{1,2} \\ \vdots \\ \theta_{(n-3),(n-2)} \end{bmatrix}. \tag{3.17}$$

The angles utilized in the computation of $d_{\theta 1}$ are derived from consecutive vectors of a directed graph, however the angle between non-consecutive vectors offer an additional structural characteristic of the graph-based region descriptor. When coupled with $d_{\theta 1}$, the descriptor would provide a rotational and scale invariant representation of a graph's structure. The angle between any two vectors, $e_k$ and $e_m$ is defined from the dot product. Using this notation, the non-consecutive vector angle descriptor is stated in (3.19).

$$\theta_{k,m} = \cos^{-1} \frac{e_k \cdot e_m}{\|e_k\| \|e_m\|} \tag{3.18}$$

$$d_{\theta 2} = \begin{bmatrix} \theta_{0,2} \\ \theta_{1,3} \\ \vdots \\ \theta_{(n-3),(n-1)} \end{bmatrix} \tag{3.19}$$

The angles from $d_{\theta 1}$ and $d_{\theta 2}$ are typically represented with floating point numbers in the range $0^\circ$ to $360^\circ$. In such a scenario, we may use the Euclidean distance for

comparing the descriptors; however we propose encoding the angles using the Gray Code encoding scheme while using the Hamming distance as a similarity measure between two descriptors. The binary encoding process starts by first converting the angle in degrees to a binary equivalent using the binary angle measurement approach. This binary string is then converted to a Gray Code string for later comparison using the Hamming distance.

### 3.3.2.1. Binary Angle Measurement

The binary angle measurement (BAM) is commonly used to represent angles as a binary string. A binary representation provides an efficient structure for data to be stored and transmitted. The primary idea behind BAM is that each bit represents $\frac{360^\circ}{2^n}$ , where $n$ denotes the number of bits in the binary string. Equation (3.21) provides the relationship between an angle, $\theta$, and its BAM representation. Fig. 3.8 offers a visual representation of the BAM system for 8-bit binary strings. The proposed angle descriptors use 8-bit binary representations.

$$BAM = \frac{2^n}{360}\theta \tag{3.21}$$

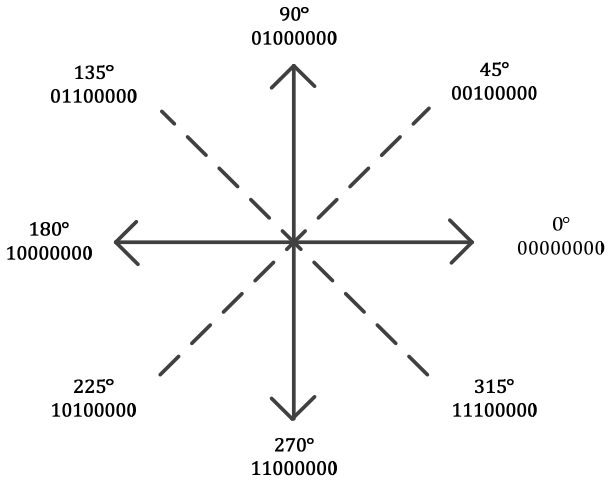



Fig. 3.8 8-bit BAM representation

### 3.3.2.2. Gray Code

Originally proposed in 1947 by Frank Gray [68], the Gray Code is a binary encoding scheme established to represent differences between consecutive values using a single bit. Table 3.1 outlines an example of a Gray Code scheme for base-10 numbers 0-7. The binary strings illustrate only a single bit change for consecutive integer changes.

Table 3.1 Gray Code encoding for base-10 range 0-7. The red indicates the single bit changed for successive value changes

| Base-10 | Binary Representation | Gray Code String |
|---|---|---|
| 0 | 000 | 000 |
| 1 | 001 | 001 |
| 2 | 010 | 011 |
| 3 | 011 | 010 |
| 4 | 100 | 110 |
| 5 | 101 | 111 |
| 6 | 110 | 101 |
| 7 | 111 | 100 |

For any given binary string, the gray code can be computed quickly. Let a binary string $b$ contain $n$ bits, $B = b_{n-1}b_{n-2} \dots b_0$, then the Gray Code binary string, $BGC = bgc_{n-1}bgc_{n-2} \dots bgc_0$ is computed according to the following,

$$bgc_i = \begin{cases} b_i & i = n-1 \\ b_i \oplus b_{i+1} & otherwise \end{cases} \tag{3.20}$$

### 3.3.2.3. Hamming Distance

The Hamming distance is an efficient method for comparing the similarity of two binary strings that are of the same length. Originally proposed by Richard Hamming [69],

the metric identifies the number of bits that differ across two binary strings. Stated formally, we define two input binary strings, $B_i$ and $B_j$, as well as the resulting binary string from the logical exclusive-OR of the input strings to be $H = B_i \oplus B_j$. The Hamming distance, $d_H$ is then defined as,

$$d_H(B_i, B_j) = \sum_{k=1}^{n} H_k, \tag{3.21}$$

where $H_k$ is bit $k$ of binary string $H$.

### 3.3.2.4. Combined Angle and Descriptor Similarity Score

The comparison of two graphs using the angle and descriptor information is completed in a similar method as depicted in Fig. 3.7 where a sub-graph of the larger graph is compared to the entire smaller graph. For each comparison, a score is generated while the smallest of the resulting scores is regarded as the similarity score for the two region descriptors. If both graphs contain the same number of nodes, a single score is generated and assigned.

When each node of the graph is represented by a SURF feature point, the similarity between two nodes is computed using $S_E$ from (3.1), whereas the angle similarity measure for two graphs, $\alpha$ and $\rho$, is defined as,

$$S_\theta = \frac{\sum_{k=1}^{|d_{\theta 1}^{\alpha}|} \left[ \frac{d_H\left(d_{\theta 1}^{\alpha}(k), d_{\theta 1}^{\rho}(k)\right)}{2^n} \right]}{\left|d_{\theta 1}^{\alpha}\right|} + \frac{\sum_{k=1}^{|d_{\theta 2}^{\alpha}|} \left[ \frac{d_H\left(d_{\theta 2}^{\alpha}(k), d_{\theta 2}^{\rho}(k)\right)}{2^n} \right]}{\left|d_{\theta 2}^{\alpha}\right|}, \tag{3.22}$$

where $d_{\theta 1}^{x}(k)$ and $d_{\theta 2}^{x}(k)$ are the $k$-th angle in the $d_{\theta 1}$and $d_{\theta 2}$descriptors, respectively, for graph $x$. The overall score for comparing two graphs is then defined as,

$$S = \frac{\sum_{|V|} S_{FP}}{|V|} + S_{\theta} \tag{3.23}$$

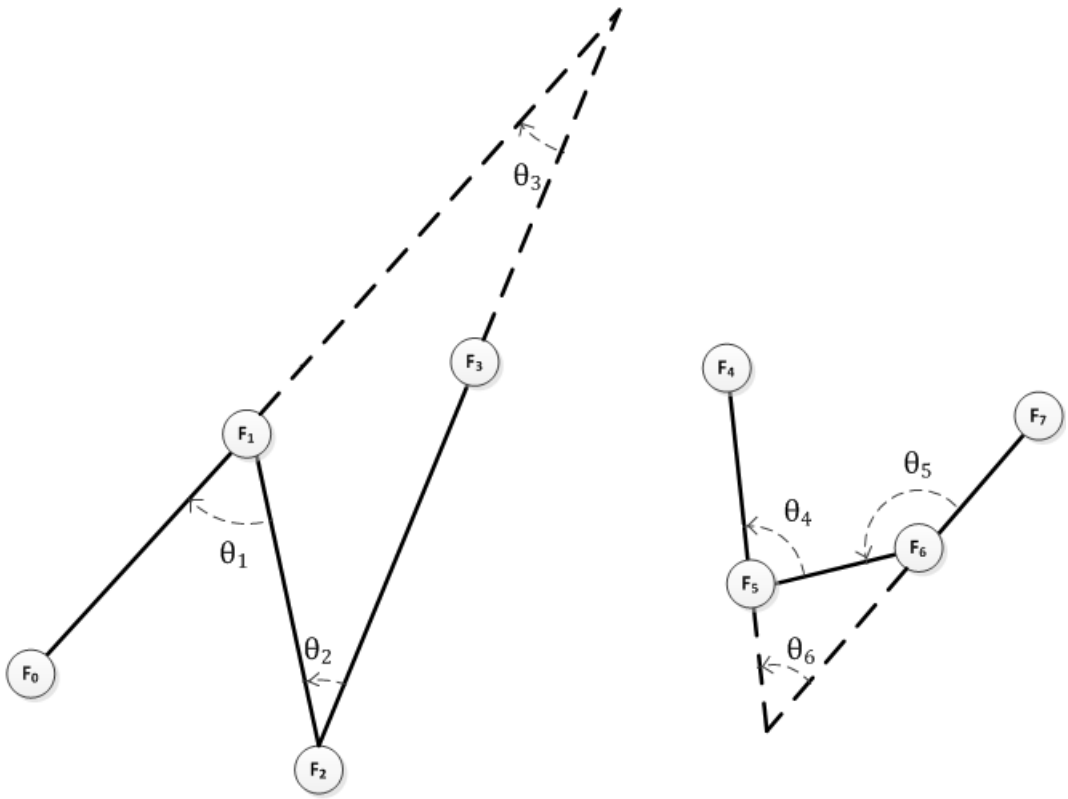



Fig. 3.9 Visual representation of the angle descriptor components

As a simple example, given two graphs of the same size, as illustrated in Fig. 3.9, we can calculate the similarity between the two graphs as,

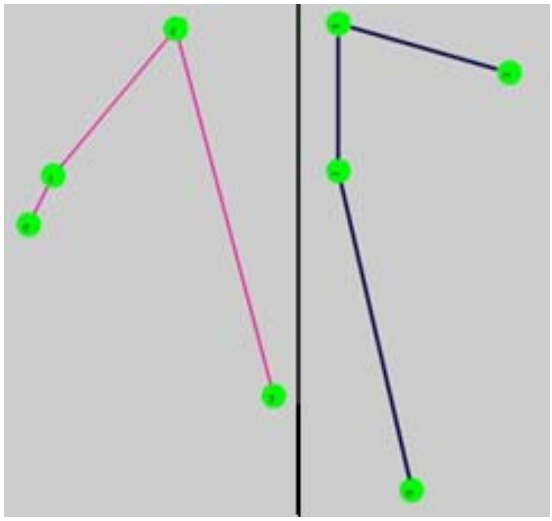
$$S = \frac{S_{FP}(F_0, F_4) + S_{FP}(F_1, F_5) + S_{FP}(F_2, F_6) + S_{FP}(F_3, F_7)}{4} + \frac{d_H(\theta_1, \theta_4) + d_H(\theta_2, \theta_5)}{2 * 2^8} + \frac{d_H(\theta_3, \theta_5)}{2^8}.$$

## 3.4. Results

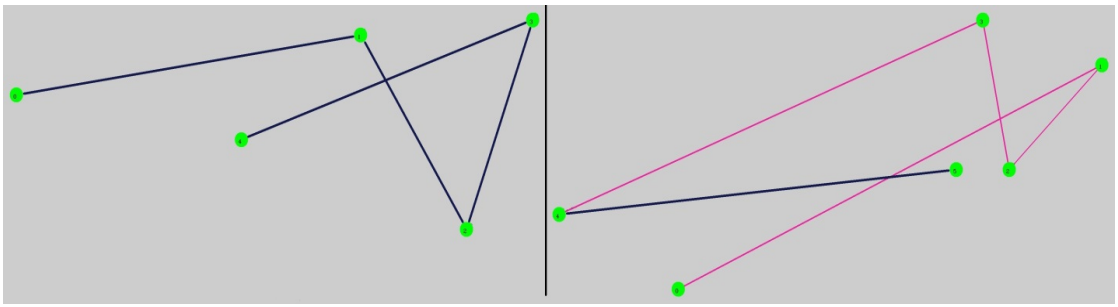
A database consisting of images from a scene before and after a disaster where used to evaluate the effectiveness of the proposed graph-based region descriptor. The test set of images are comprised of images from urban landscapes, aerial photography, scenes before and after a disaster, as well as a few general objects to demonstrate the robust matching of the proposed descriptor.

### 3.4.1. Graph Matching

The proposed region descriptor matching scheme is shown to be effective for identifying graphs of similar structure, as shown in Fig. 3.10. The descriptors defined in (3.17) and (3.19), coupled with the similarity metric in (3.22), present a rotational and scale invariant approach for comparing the structure of two graphs, whereas the contribution to the overall similarity score in (3.23) from $S_{FP}$ ensures pixel distributions between the two graphs exhibit related textural characteristics.



a. Two matched graphs with equal number of nodes



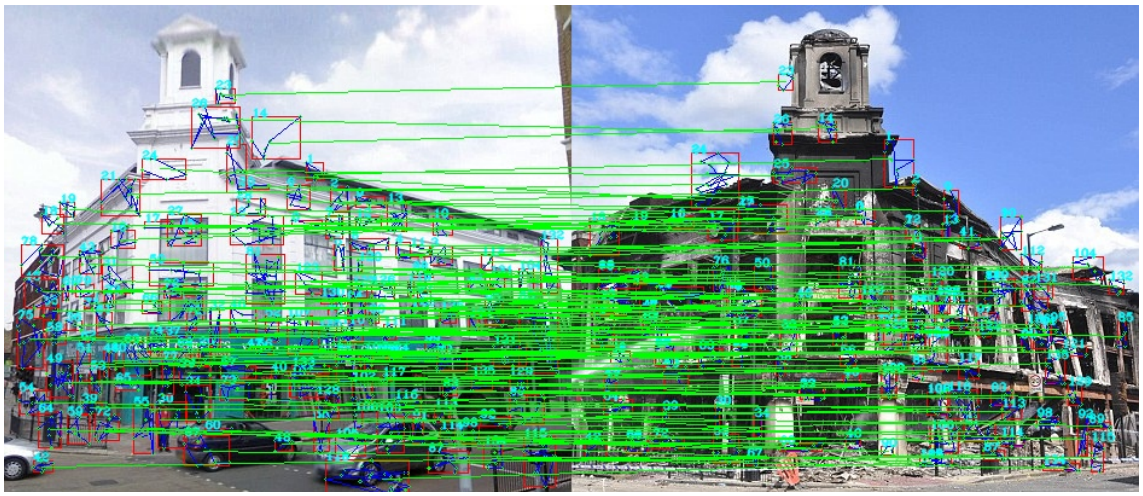
b. Illustrates an example of graphs with differing number of nodes. The red highlighted sub-graph on the right was found to be the best match for the graph on the left

Fig. 3.10 Two examples of matched graphs using SURF feature points and their angles

### 3.4.2. Feature Matching Across Images

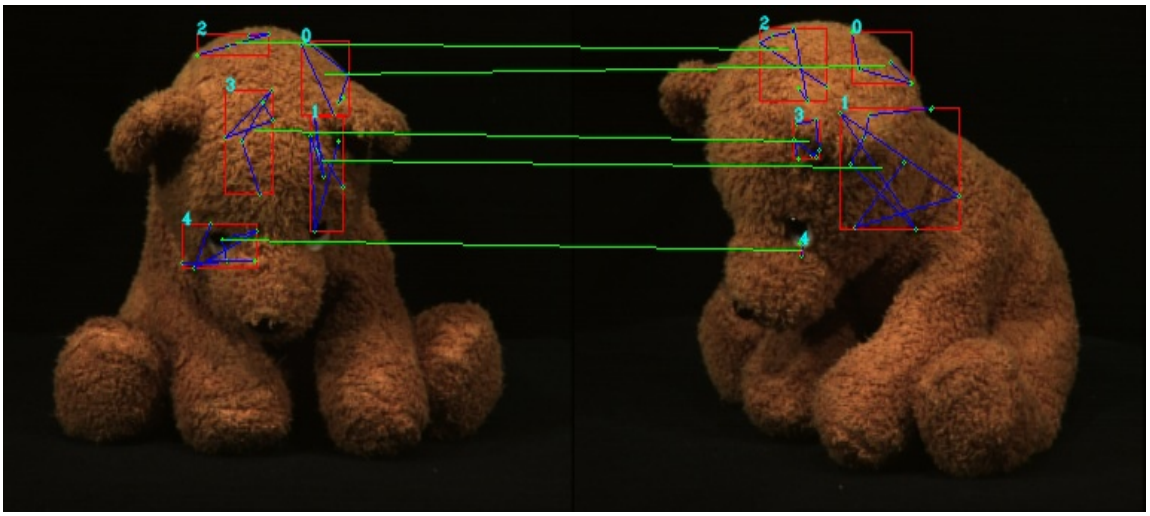
For all three scenarios evaluated, the proposed region descriptor demonstrated robust ability for feature matching in urban scenes, aerial photographs and when applied to general objects. The building and aerial images are greatly affected by both photometric and geometric variations due to the state of the scene after a disaster, whereas the general objects exhibit little photometric difference due to lighting variations but show the objects from different perspectives.

The images in Fig. 3.11 demonstrate the robust nature of the proposed graph-based region descriptor for feature matching. Fig. 3.11a and Fig. 3.11b depict scenes after and during a disaster, respectively. Both scenarios exhibit a high number of correctly matched features even under local deformations that range from content differences to geometric variations. The examples in Fig. 3.11c and Fig. 3.11d provide strong evidence that the proposed region descriptor is effective for general objects from varying perspectives. Lastly, the proposed method is shown to match feature accurately for aerial photos of natural landscapes in Fig. 3.11e.

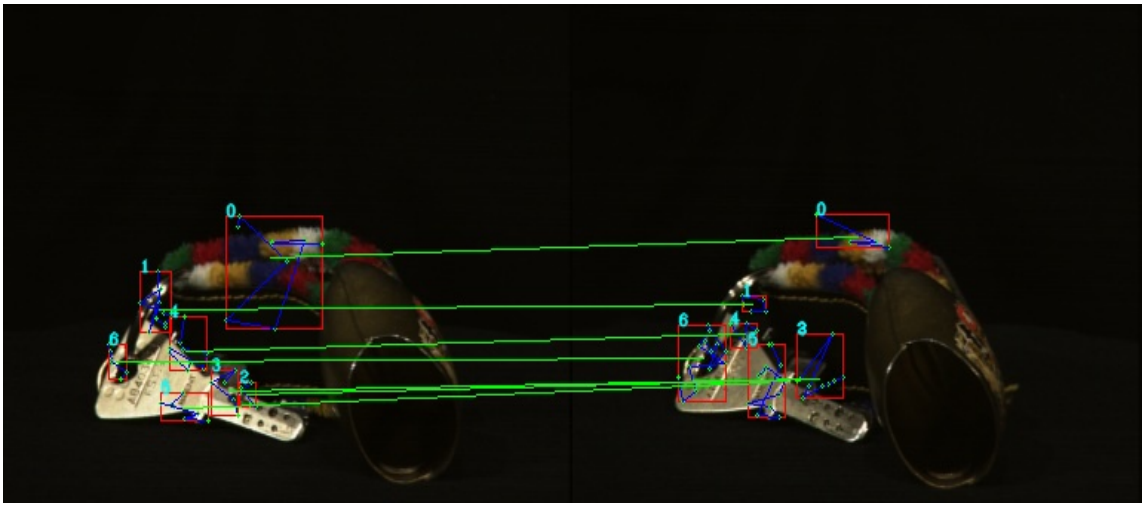


a. Descriptor's matching abilities for urban landscapes after a disaster

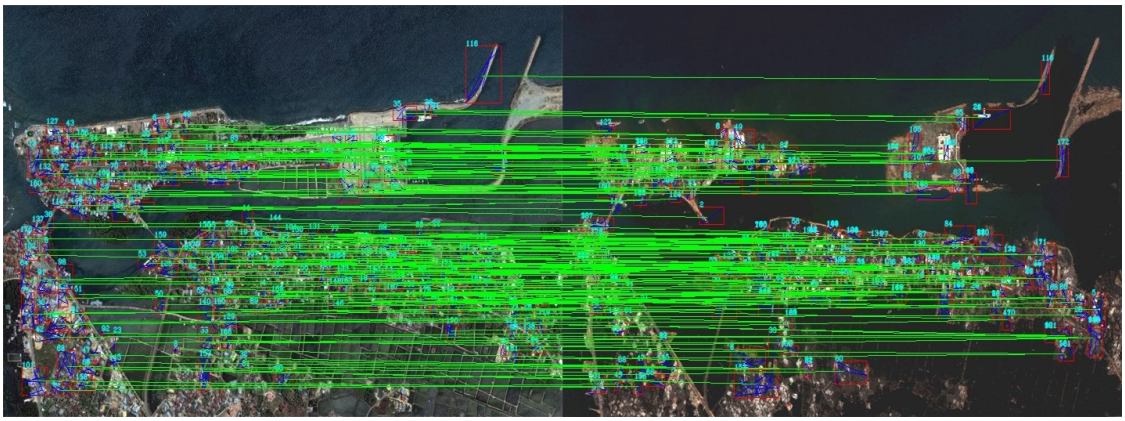
b. Descriptor's matching abilities for urban landscapes during a disaster



c. Example descriptor use with general objects that are affected by perspective variations



d. Example descriptor use with general objects that are affected by perspective variation and illumination differences



e. Depicts the proposed region descriptor utilized with an aerial image of a territory affected by natural disaster

Fig. 3.11 Example image pairs demonstrating the feature matching abilities of the proposed region descriptor

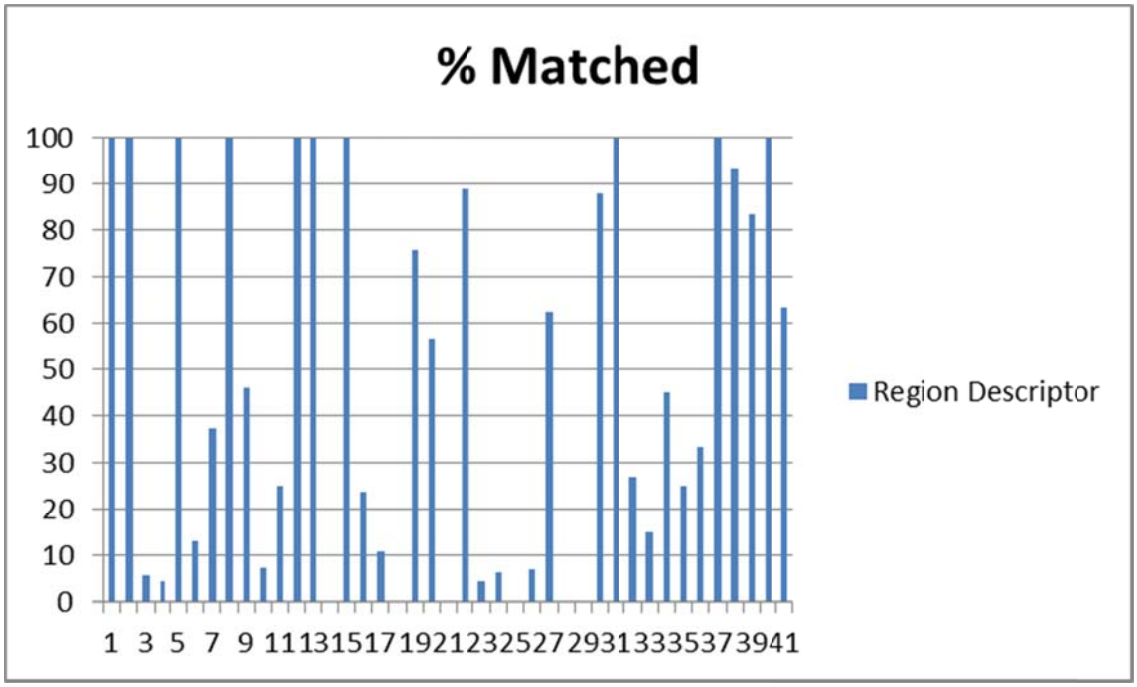
### 3.4.3. SIFT, SURF and BRISK Matching Comparisons

Given two images, $I_1$ and $I_2$, that vary under a projective transform, there exists a transformation matrix, $H$, such that $I_2 = HI_1$. If the transformation matrix is known, the location of a feature point in $I_1$ can be identified in $I_2$ according to, $\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = H \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$, where $(x_1, y_1)$ and $(x_2, y_2)$ are the spatial coordinates of feature points in $I_1$ and $I_2$, respectively. The following process was used to identify the number of correctly matched feature points for a group of 41 matched images.

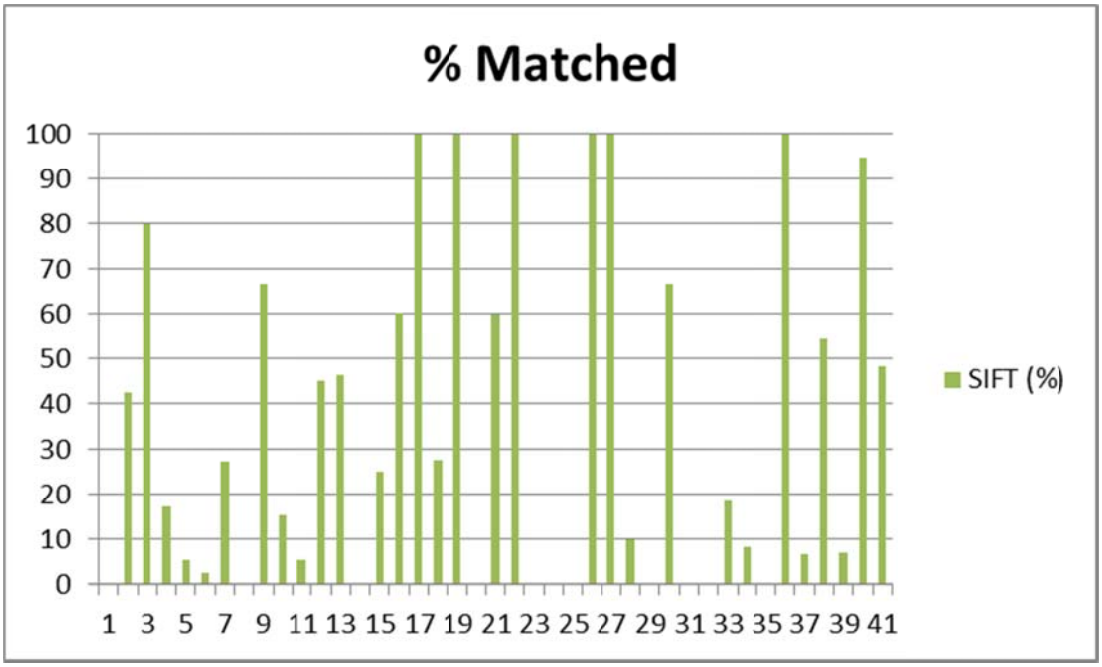
- For each image pair, manually select 10 matching control points across both images.
- Estimate the transformation matrix using the Direct Linear Transform.
- For each feature point in the reference image, calculate the projected point in the query image using the estimated transformation matrix.

- Calculate the spatial distance between the projected point and the matched feature point of the query image. If the two points are within 10 pixels, count the pair as a match.

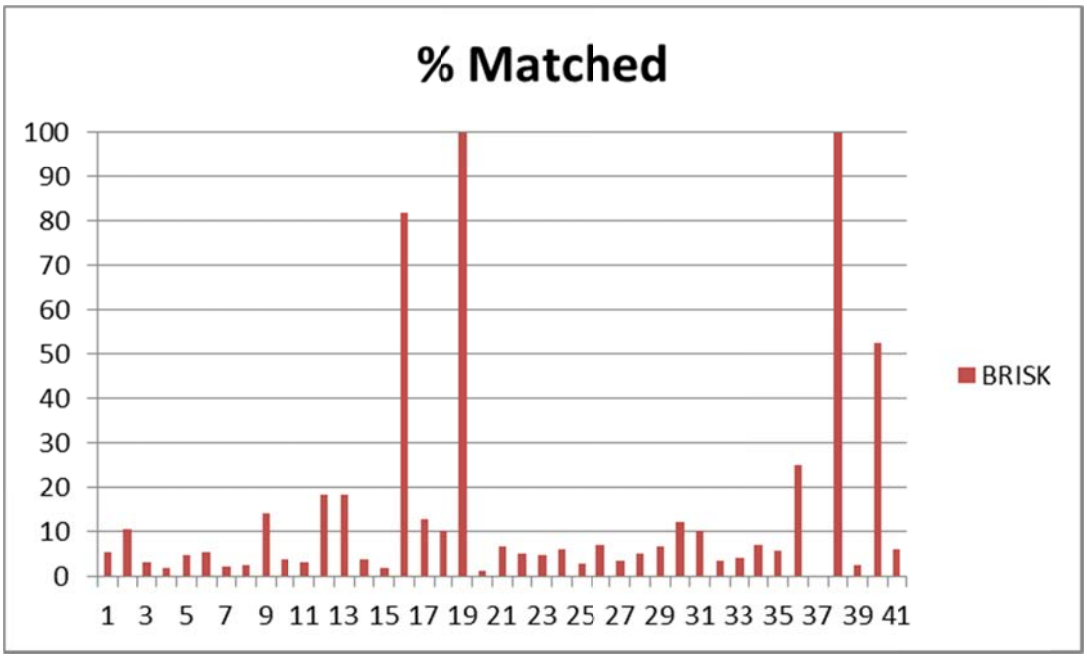
The following graphs in Fig. 3.12 present the number of matched feature points for each image pair. It is shown that the proposed region descriptor performed better in most scenarios than SIFT, SURF and BRISK.



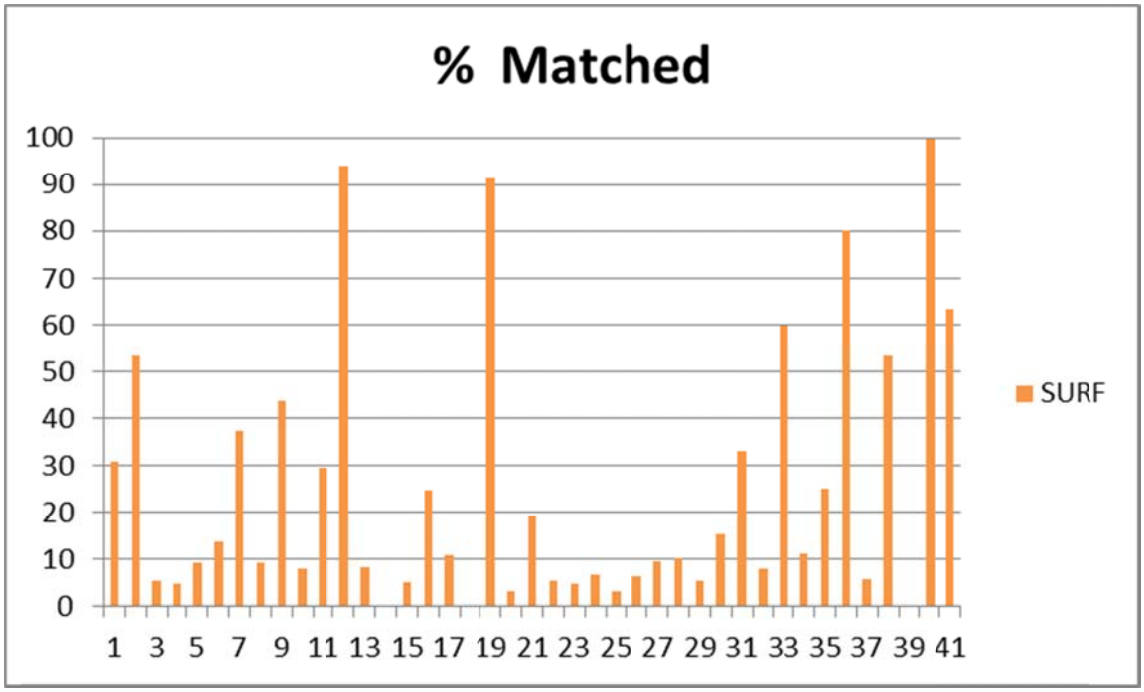

a. Percent correctly matched proposed region descriptors for 41 image sets




b. Percent correctly matched SIFT descriptors for 41 image sets




c. Percent correctly matched BRISK descriptors for 41 image sets




d. Percent correctly matched SURF descriptors for 41 image sets

Fig. 3.12 Graphs illustrating the percent of feature points matched between 41 sets of image pairs

The following graph in Fig. 3.13 compares the matching accuracy distribution of the proposed region descriptor to SURF, SIFT and BRISK distributions. As shown, BRISK demonstrated the poorest performance, as depicted by the majority of test cases exhibiting less than 20% matching accuracy. For the aerial and street view images, SIFT outperformed SURF, while the proposed region descriptor is observed to have a higher matching accuracy where most test cases demonstrated a matching accuracy of at least 60%. The distribution parameters are summarized in Table 3.2, where it is evident that the proposed region descriptor provides a higher average matching rate than the traditional feature descriptors. As shown in Table 3.2, the graph-based region descriptor produced an average matching accuracy of 62.19%, where traditional approaches using SURF, SIFT and BRISK had matching accuracies of 24.65%, 35.18%, and 14.2%,

respectively, which indicates the effectiveness of the graph-based region descriptor in situations involving disaster scenes.

The overall distributions depicted in Fig. 3.13 Matching accuracy distribution for the proposed region descriptor, SURF, SIFT and BRISK. indicate the effectiveness of the proposed region descriptor by illustrating the descriptor's superior distribution for higher matching rates. Similarly, for lower matching rates, the proposed descriptor is observed to have the lowest distribution, further indicating the proposed method's robustness in situations that involve large variations which adversely affect the traditional approaches. Each of the image sets can be categorized according to the difficulty associated with feature matching. The sources of such difficulties may be a combination of geometric variations from perspective differences, and image content variations from the effects of a disaster, object occlusion, or lighting conditions. 31.7% of the test cases exhibit examples of each difference that adversely affect the matching process. In these instances, the matching rate is low; however the proposed region descriptor still outperforms the traditional methods. Similarly, approximately 24.6% of the test cases provide examples where the variations are not as extreme and therefore the matching results are more favorable. Lastly, approximately 43.7% of the test data is comprised of images where distinct features are easily matched across images. In all three scenarios, the proposed region descriptor is shown to be more robust than the traditional methods.

Table 3.2 Matching accuracy distribution parameters.

| **Matching Accuracy (%)** | | | | |
|---|---|---|---|---|
| | **SURF** | **SIFT** | **BRISK** | **Region Desc (Proposed Method)** |
| **Average** | 24.65 | 35.18 | 14.20 | **62.19** |
| **Std Dev** | 27.72 | 36.77 | 24.15 | **35.05** |

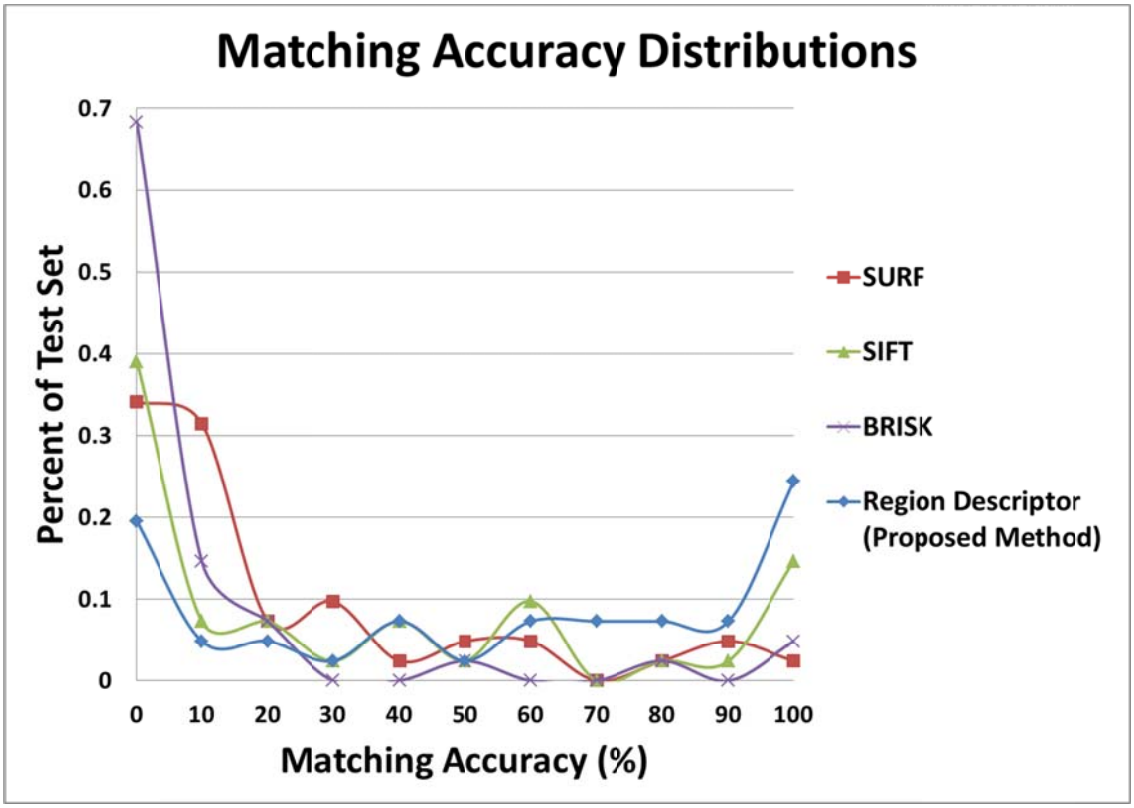



Fig. 3.13 Matching accuracy distribution for the proposed region descriptor, SURF, SIFT and BRISK.

### 3.5. Summary

In this chapter, a novel graph-based region descriptor is proposed. The descriptor couples invariant feature point descriptors with the spatial relationship to neighboring feature points. After the initial detection of feature points, clustering is performed to group similar keypoints. Once clustered, the graphs are constructed using a shortest-path process. Additionally, an effective similarity measure is provided for matching the proposed region descriptor. It has been shown to be more effective than matching individual SURF feature points in the scenario of urban landscapes and aerial imagery.

# 4. COMPREHENSIVE FEATURE AND TEXTURE FUSION-BASED IMAGE REGISTRATION APPROACH

Image-based situational awareness may require images of the same scene, event or object, but from differing perspectives, to be analyzed from a mutual coordinate system. Furthermore, fusion of such data is relevant in applications where partial information from multiple sources can offer a comprehensive perspective of a scene, situation, or event. Applications including disaster scene analysis, emergency assessment, early warning systems, and statistics collection, can benefit from a universal view of a given data set.

Registration is a common task that estimates the transformation parameters relating two images. The aim is to project one image such that both images contain some region which overlap and may appear to be from the same perspective. This region of interest isolates areas of the images that may share common details and features. Moreover, areas designated as not mutual may offer desired data for a particular application, such as disaster risk management. The estimation of the transformation parameters can be accomplished once a suitable set of matched points or features are identified that accurately represent the geometric differences. Once the homography is computed, a similarity measure is used to gauge the effectiveness of the overall process.

The proposed algorithm relies on a coarse search to identify similar regions between images. To complete this task, pixel intensities are compared directly using the Normalized Cross Correlation between localized image patches. With the identified similar regions, a fine search is executed such that invariant feature points are extracted as candidate control points. The set of candidate points are further reduced using a

forward-backward matching process. Finally, the remaining control points are used in a variation of the DLT algorithm for estimating the homography. Fig. 4.1 provides an overview of the main phases of the algorithm. Using the estimated homography parameters, the target image is remapped using bilinear interpolation to minimize aliasing.

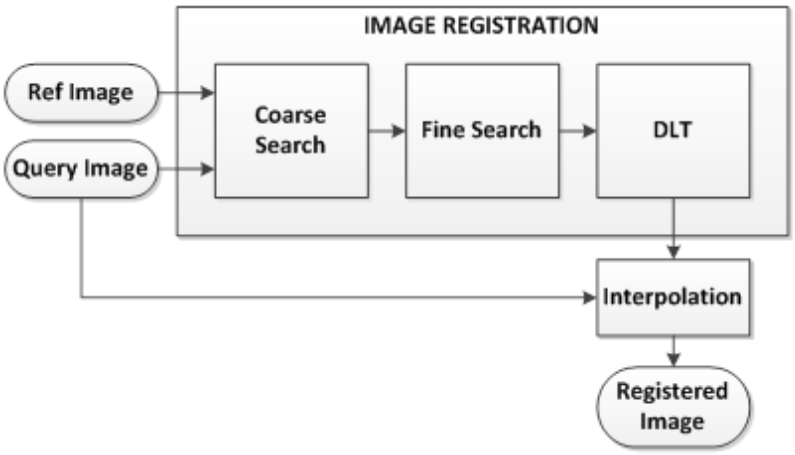



Fig. 4.1 Algorithm overview

### 4.1. Coarse Image Search

Given reference and target images, the query image is segmented into equal sized blocks, as shown in Fig. 4.2.

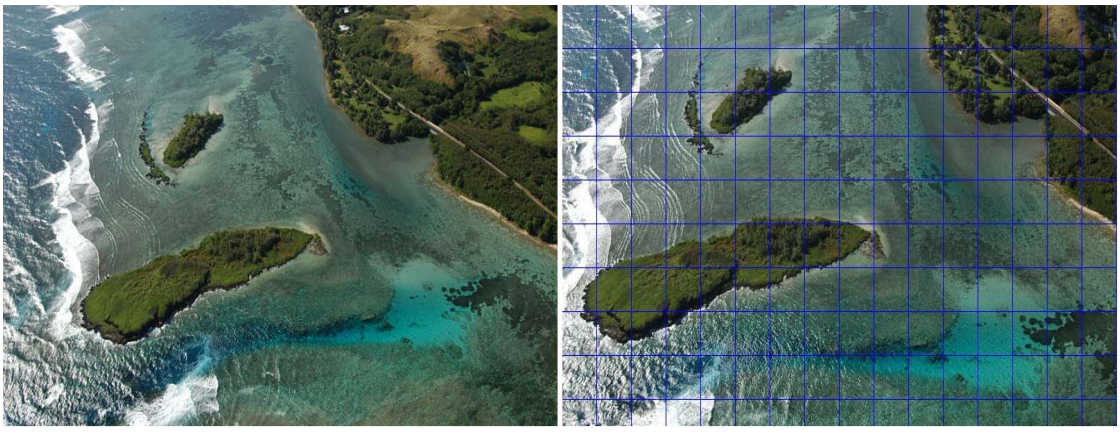


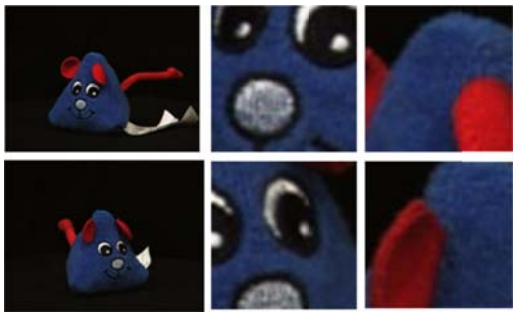
Fig. 4.2 Reference image (left) and segmented query image (right)

Each block of the target image is compared to a sliding window of the reference image. A comparison value is determined using the Normalized Cross Correlation (NCC).

$$NCC = \frac{1}{xy-1}\sum_{i=1}^{x}\sum_{j=1}^{y}\frac{\left(f(i,j)-\bar{f}\right)(t(i,j)-\bar{t})}{\sigma_f \sigma_t} \tag{4.1}$$

where x and y are the dimensions of the image patches while $\bar{f}$ and $\bar{t}$ are the average value of all pixels within the reference and query patches, respectively. The most similar patches maximize the NCC metric. Fig. 4.3 provides example similar patches identified using the NCC course search.

a. Similar patches extracted from an aerial photograph, where patches from the reference image span the top row and patches from the query image along the bottom row



b. Matched patches extracted from the images in the left-most column

Fig. 4.3 Similar patches identified using the normalized cross-correlation

## 4.2. Fine Image Search

After the identification of matched image patches, the proposed method detects, extracts and matches invariant feature points which will be the basis for the transformation matrix estimation. The coarse search reduces the search space for a set of feature points. For this work, SURF keypoints are exploited with a symmetric matching scheme for improved matching accuracy. A detailed overview of SURF can be found in Section 2.2.1.2.

The similarity between two SURF points is simply the Euclidean distance of their 64-dimensional descriptor (4.2); however, for objects with a repetitive structure, such as buildings with many similar windows, the SURF descriptor is not discriminative enough to provide a one-to-one matching across images.

$$ED = \sqrt{\sum_{i=1}^{64} \left(d_1^k[i] - d_2^m[i]\right)^2} \tag{4.2}$$

where $d_1^k[i]$ and $d_2^m[i]$ represent the $i$-th component of descriptors $k$ and $m$, from image 1 and 2, respectively.

If $ED_1^{k,m}$ is the Euclidean distance of the best match for descriptor $k$ from image 1, and $ED_1^{k,n}$ is the distance of the second best match for descriptor $k$, then the matched descriptor pair is added to the set according to,

$$MP_{12} = MP_{12} \cup \left\{(k,m) \mid \frac{ED_1^{k,m}}{ED_1^{k,n}} < 0.85\right\} \tag{4.3}$$

The set $MP_{12}$ represents the set of matched pairs that satisfy (4-3) using image 1 as the reference and image 2 as the target image. Similarly, a second set, $MP_{21}$ is also formulated but using image 2 as the reference and image 1 as the target image.

$$MP_{21} = MP_{21} \cup \left\{(j,n) \mid \frac{ED_2^{j,n}}{ED_2^{r,n}} < 0.85\right\} \tag{4.4}$$

where $ED_2^{j,n}$ denotes the smallest distance between descriptor $n$ from image 2 and $j$ of image 1, while $ED_2^{r,n}$ is the distance for the second best match for descriptor $n$. A third set is constructed as the intersection of the first two.

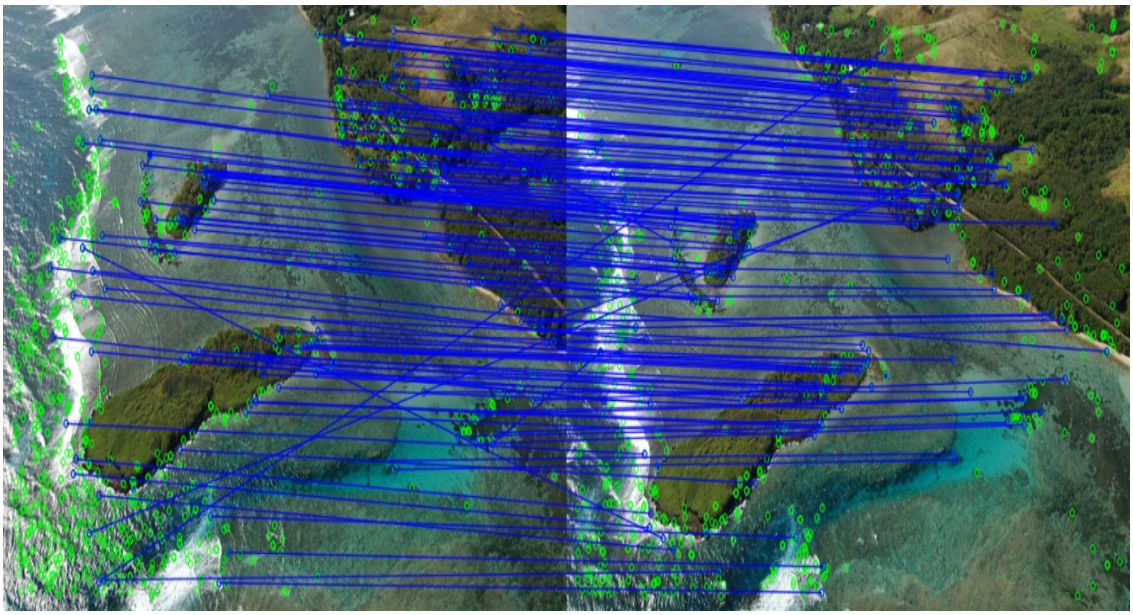
$$MP_S = MP_{12} \cap MP_{21} \tag{4.5}$$

$MP_S$ represents the set of symmetric match pairs. In other words, the best matched pair from image 1 to 2 is also the best matched pair from image 2 to 1. This condition ensures a one-to-one mapping of local feature points. To further improve the accuracy of the SURF matching process, the set of symmetric matches is examined to determine the best matched pair. Let $d_{BM}$ denote the best matched distance between two feature points in $MP_S$. Then the set $MP_S$ is pruned to eliminate any matched pairs according to the following, which results in a new set, $MP_S^*$,

$$MP_S^* = \{MP_S | ED \leq 2 * d_{BM}\}, \tag{4.6}$$

where $ED$ denotes the set of matched distances for each feature point pair in $MP_S$. The effects of symmetric matching are exemplified in Fig. 4.4.

a. SURF matches using FLANN matcher without symmetric matching



b. SURF using FLANN matcher and symmetric matching. As illustrated, many outlier matches have been omitted

Fig. 4.4 Example SURF matching with and without symmetric matching

### 4.3. Transformation Parameter Estimation

Image registration is a process that aims to map one image to another through an estimation of a transformation matrix. To do so, at least four non-collinear matched control points are required in each image. Let $(x_i, y_i)$ be the coordinates in the reference image, and $(x_i', y_i')$ be the control points in the query image, where $i = 1,2,3,4$. Then visually, the goal is to find the transformation matrix which maps one shape to the other as illustrated in Fig. 4.5.

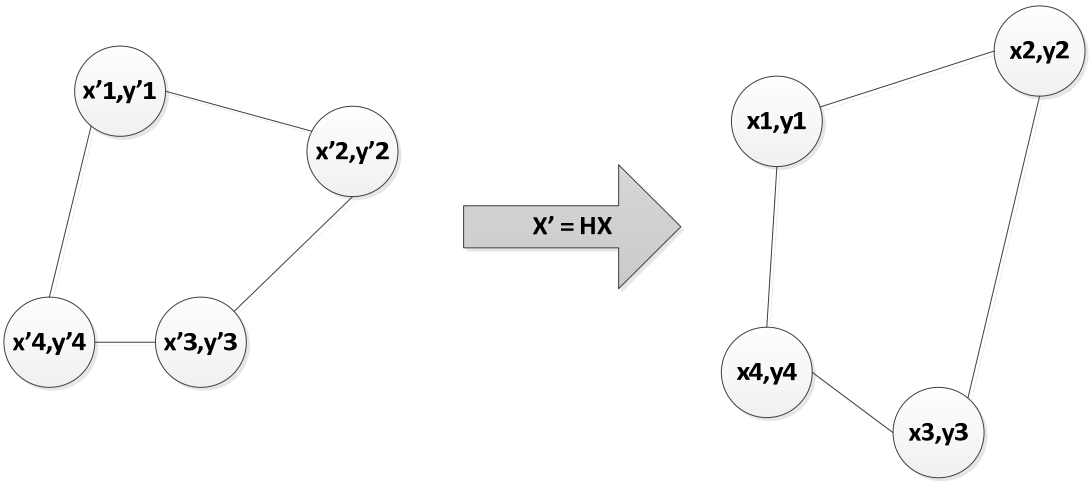



Fig. 4.5 Control point matching

The Direct Linear Transform (DLT) algorithm outlined in [70] provides a novel approach for estimating the homography matrix through the Singular Value Decomposition. Let $\boldsymbol{x} = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$ and $\boldsymbol{x}' = \begin{bmatrix} x_1' \\ y_1' \\ z_1' \end{bmatrix}$ be the normalized control point pairs such that $\boldsymbol{x}' = \boldsymbol{H}\boldsymbol{x}$. Using the normalized control points, a matrix $A \in \mathbb{R}^{3n\,x\,9}$ is constructed such that,

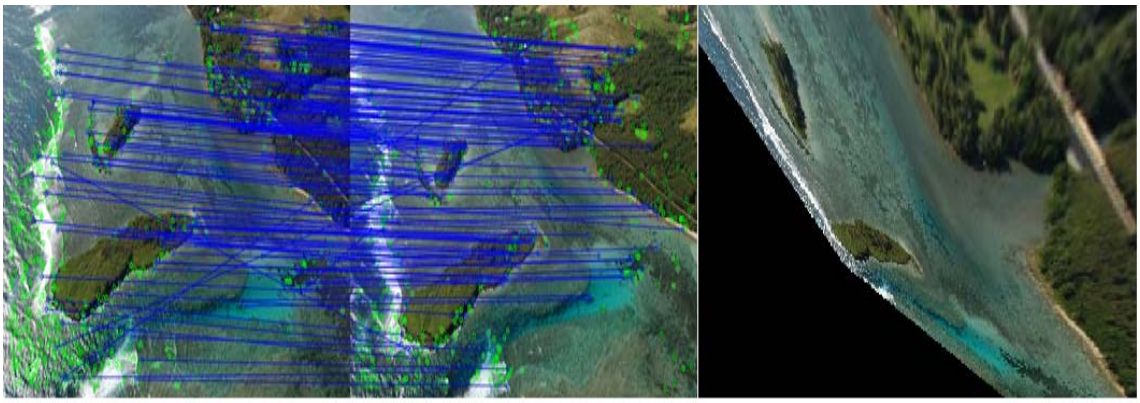
$$A_i = \begin{bmatrix} 0 & 0 & 0 & -x_i z_i' & -y_i z_i' & -z_i z_i' & x_i y_i' & y_i y_i' & z_i y_i' \\ x_i z_i' & y_i z_i' & z_i z_i' & 0 & 0 & 0 & -x_i x_i' & -y_i x_i' & -z_i x_i' \\ -x_i y_i' & -y_i y_i' & -z_i y_i' & x_i x_i' & y_i x_i' & z_i x_i' & 0 & 0 & 0 \end{bmatrix} \quad (4.7)$$

$$A = \begin{bmatrix} A_1 \\ \vdots \\ A_i \end{bmatrix}, i = 1 \dots n \tag{4.8}$$

The singular value decomposition of $A$ yields the homography parameters in the last column of the right singular vector matrix, $V$.

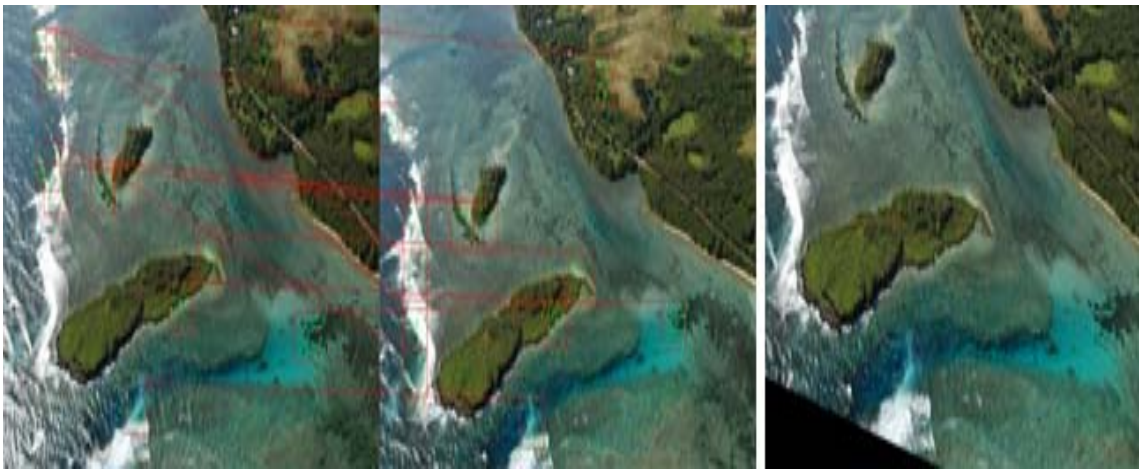
### 4.4. Results

As a baseline for comparison, the image registration technique used in [71] was implemented. These works relied on SURF feature points and the Random Sample Consensus (RANSAC) method for the detection of matched control points. Fig. 4.6 illustrates an example of the registration results strictly using the best matched SURF feature points and DLT. Visual inspection of the results verifies the inaccuracies of the matched control points. Fig. 4.7 provides results using the same set of images but utilizing the proposed algorithm.



a. b. c.

Fig. 4.6 Matched SURF feature points from a reference (a) and query image (b). The resulting registration utilizing the SURF points and the DLT method for transformation estimation (c)



a. b. c.

Fig. 4.7 Reference and query images shown with similar patches denoted in red boxes and identified control points designated with green circles (a). The remapped query image after registration (c)

The remainder of the tests was conducted on images from the Amsterdam Library of Object Images [72], which provides many object images taken from different perspectives and lighting conditions. The average computation time for the proposed algorithm was 5 seconds for each image of resolution 320x240.

In instances where six or more correct control points are identified, the effect of outlier control points on the registration accuracy was negligible, as is illustrated in Fig. 4.8. With the inclusion of more correctly matched control points, the parameter estimation becomes more resilient to the incorrectly identified points. Through the two-phase search algorithm involving the coarse image search, followed by the fine search through invariant feature detection, the proposed method has demonstrated high accuracy in identifying correctly matched control points.

a. e. i.


b. f. j.

c. g. k.

d. h. l.

Fig. 4.8 Registration results depicting the reference images (a-d), query images (e-h), and transformed query image (i-l). The red boxes indicate the matched image patch while the green circles denote the detected feature points

As realized in Fig. 4.8, the proposed hybrid method can be effective in scenarios involving high contrast, such as the examples shown in Fig. 4.8a and Fig. 4.8b. The example in Fig. 4.8c suggests the method is effective for aerial imaging where small geometric variations exist such as translations and rotations. Applications involving image stitching and mosaicking of natural scenes may benefit from the hybrid technique. The example illustrated in Fig. 4.8c provides an example where the proposed method could use improvement. Such situations are characterized by objects with repeated structural elements. In these scenarios, the coarse template matching will be adversely affected by the image content. Since the overall registration is dependent upon the accuracy of the coarse search, the overall registration will be inaccurate, as shown by the two mismatched image areas depicted in Fig. 4.8d and Fig. 4.8h.

## 4.5. Summary

In this chapter a novel image registration technique that employs both intensity and feature-based analysis for the identification of suitable control points across images is proposed. Initial image region similarities are identified through an exhaustive search involving the Normalized Cross Correlation measure. By determining the image patches with maximized NCC values, one is able to drastically narrow the feature search space.

The next phase involves computation of invariant features that are matched using a two-phase matching process. This forward and backward matching ensures one-to-one correspondence of control points and minimizes the amount of potential outlier points. Lastly, the Direct Linear Transform is utilized to estimate the homography between images for the over-determined system. Because the proposed solution does not rely on optimization algorithms, the results are not affected by the discovery of a global optimum, which may negatively affect the common approaches. The results verify the accuracy of the proposed method through the efficient calculation of correctly matched control points.

# 5. MULTISTAGE FEATURE-BASED IMAGE REGISTRATION

Images depicting a scene before and after a disaster pose many challenging problems for image processing algorithms such as image registration. Due to the potentially large variations in textural and geometric properties, a single image registration algorithm cannot be solely used for all test cases. To accommodate a larger range of scenarios, we propose a multi-stage method that evaluates the results along each phase to determine if subsequent stages are necessary. The aim of such an approach is to broaden the image registration capabilities for images under large amounts of photometric and geometric variation.

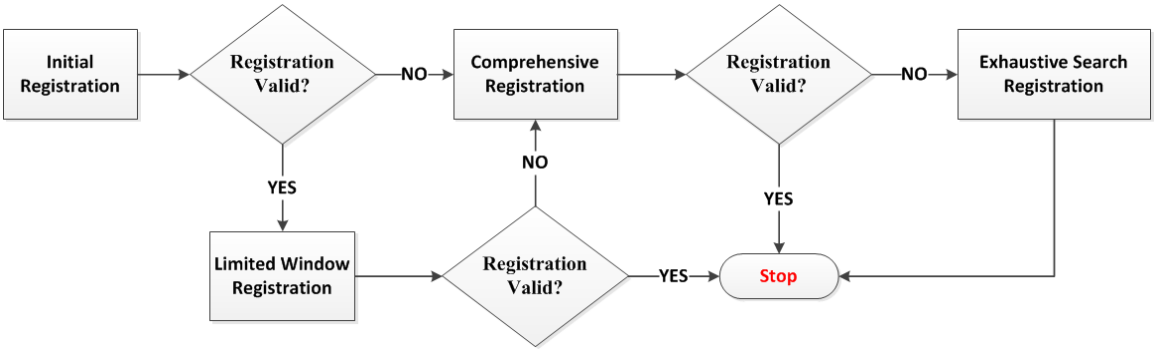



Fig. 5.1 General overview of the tasks involved in the proposed image registration method

The initial step is to determine a coarse registration exploiting the translational properties of the Fourier Transform. This approach is a common preprocessing task; however we propose a method for recovering the rotation using image edge maps instead of the typical grayscale images. The coarse registration phase is then evaluated using a simple approach involving the analysis of color pixel distributions. Per the registration verification step, an appropriate registration method is performed. The verification stage

is then revisited to determine if the current output is sufficient or whether another method should be attempted. An output is guaranteed by the end of the third attempt, as illustrated in the broad overview in Fig. 5.1 and a more detailed summary in Fig. 5.2.

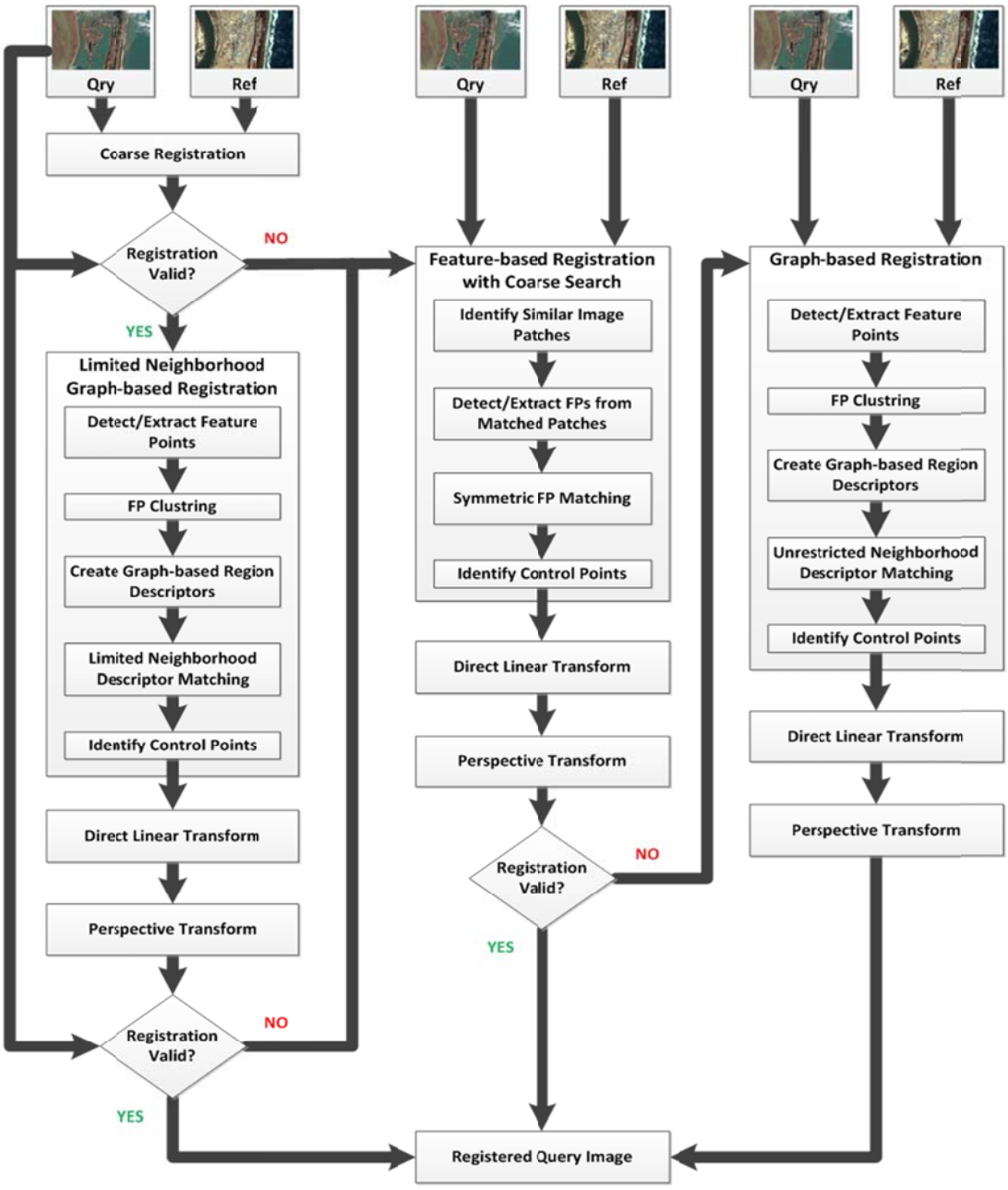



Fig. 5.2 Detailed overview of proposed registration system

## 5.1. Coarse Registration

An initial coarse registration is attempted to improve feature matching and the subsequent registration. Fig. 5.3 provides a summary of the registration process that includes the calculation of translation and rotational relationships between two images.

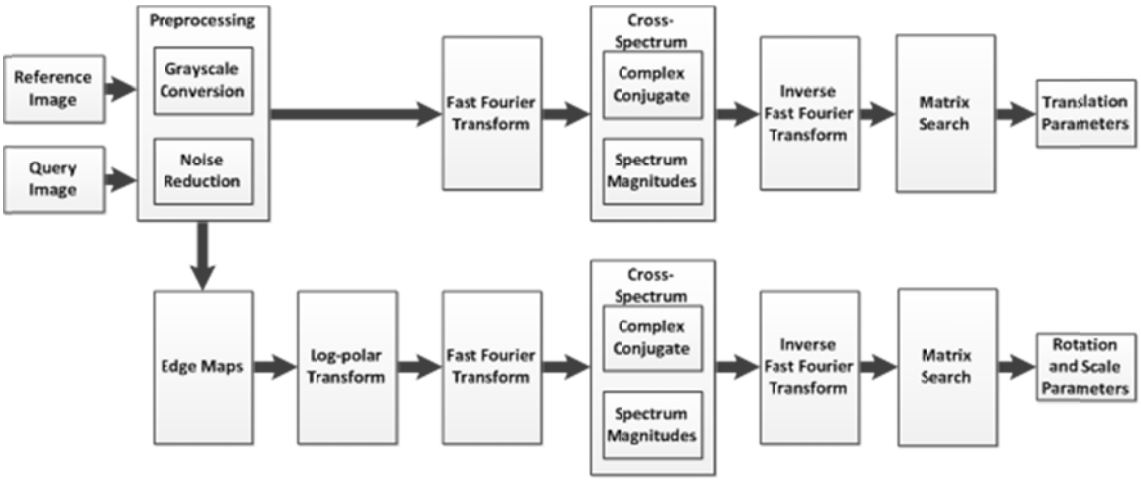



Fig. 5.3 Proposed Coarse Registration

In order to perform coarse registration, the images must first be converted from the RGB color space to grayscale. There are many approaches for this common task, however the approach used in this work simply disregards the chrominance components after converting the color space to YUV (luminance-chrominance) color space, which uses difference weights to accurately model human perception. The luminance component, $Y$, is expressed in RGB components, $R$, $G$, and $B$, according to

$$Y = 0.299R + 0.587G + 0.114B. \tag{5.1}$$

Once represented in grayscale, both input images are filtered for noise. A simple 3x3 Gaussian filter is used to smooth the images by eliminating noise. In the time domain, the images are convolved with the Gaussian kernel which is defined as,

$$G(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x}+\frac{y^2}{\sigma_y}\right)}. \tag{5-2}$$

Here, $\sigma_x$ and $\sigma_y$ are the standard deviations for the $x$ and $y$ components and dictate the bandwidth of the filter. For the proposed registration method, $\sigma_x$ and $\sigma_y$ are defined as 2 pixels.

### 5.1.1. Translation Estimation

The Fourier Shift Theorem stated in (5.3) offers a valuable tool for determining the translation, rotation, and scale parameters that relate two images. Let $\mathcal{F}\{\cdot\}$ denote the Fourier transform for signal $f(t)$, then the theorem states,

$$\mathcal{F}\{f(t-\alpha)\} = e^{-j2\pi s\alpha}F(s), \tag{5.3}$$

where $\alpha$ and $s$ represent the offset and frequency domain variable.

If two images differ only by a translation, $I_1(x,y) = I_2(x-x_t, y-y_t)$, then the image representations in the frequency domain is given as,

$$I_1(u,v) = e^{-j2\pi(ux_t+vy_t)}I_2(u,v), \tag{5.4}$$

where $I_1(u,v) = \mathcal{F}\{I_1(x,y)\}$ and $I_2(u,v) = \mathcal{F}\{I_2(x,y)\}$. Equation (5.4) can be expressed as,

$$\frac{I_1(u,v)}{|I_1(u,v)|} = e^{j2\pi(ux_t+vy_t)}\frac{I_2(u,v)}{|I_2(u,v)|} \tag{5.5}$$

$$\frac{I_1(u,v)|I_2(u,v)|}{|I_1(u,v)|I_2(u,v)} = e^{j2\pi(ux_t+vy_t)} \tag{5.6}$$

$$\frac{I_1(u,v)I_2^*(u,v)}{|I_1(u,v)I_2(u,v)|} = e^{j2\pi(ux_t+vy_t)}, \tag{5.7}$$

where $I_2^*(u,v)$ is the complex conjugate of $I_2$. The normalized cross-spectrum, $\frac{I_1(u,v)I_2^*(u,v)}{|I_1(u,v)I_2(u,v)|}$ in (5.7) is an impulse function in the time domain at the location of the translation parameters $(x_t, y_t)$.

$$\mathcal{F}^{-1}\left\{\frac{I_1(u,v)I_2^*(u,v)}{|I_1(u,v)I_2^*(u,v)|}\right\} = \mathcal{F}^{-1}\left\{e^{j2\pi(ux_t+vy_t)}\right\} = \delta(x-x_t, y-y_t) \tag{5.8}$$

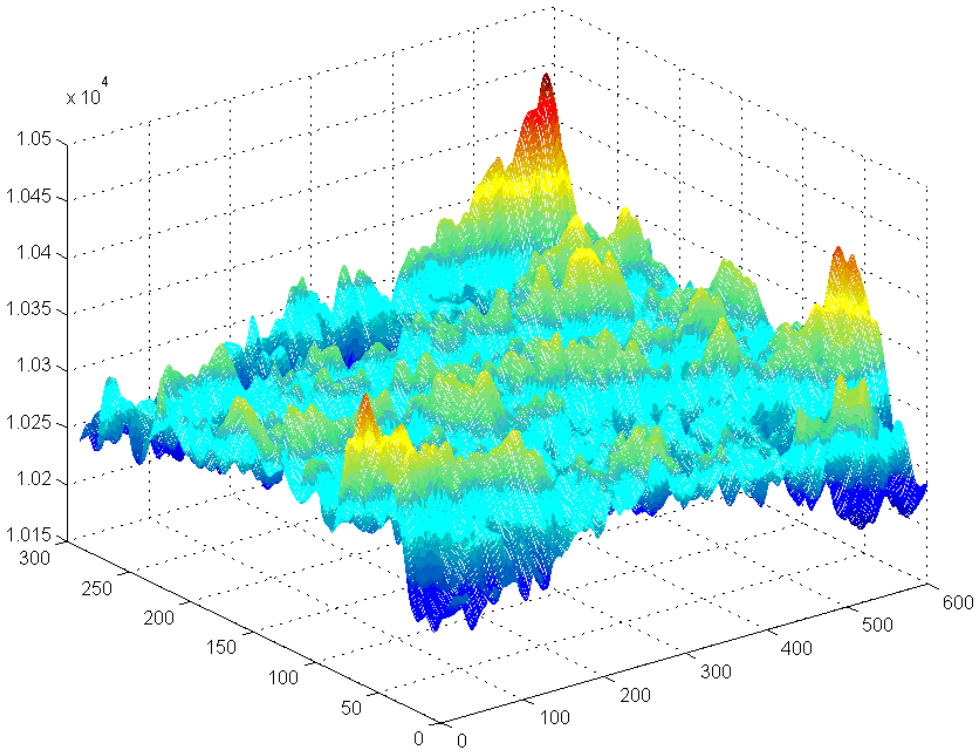
Therefore the registration parameters may be recovered according to (5.9).

$$(x_p, y_p) = \max_{x,y}\left[\mathcal{F}^{-1}\left\{\frac{I_1(u,v)I_2^*(u,v)}{|I_1(u,v)I_2^*(u,v)|}\right\}\right] \tag{5.9}$$

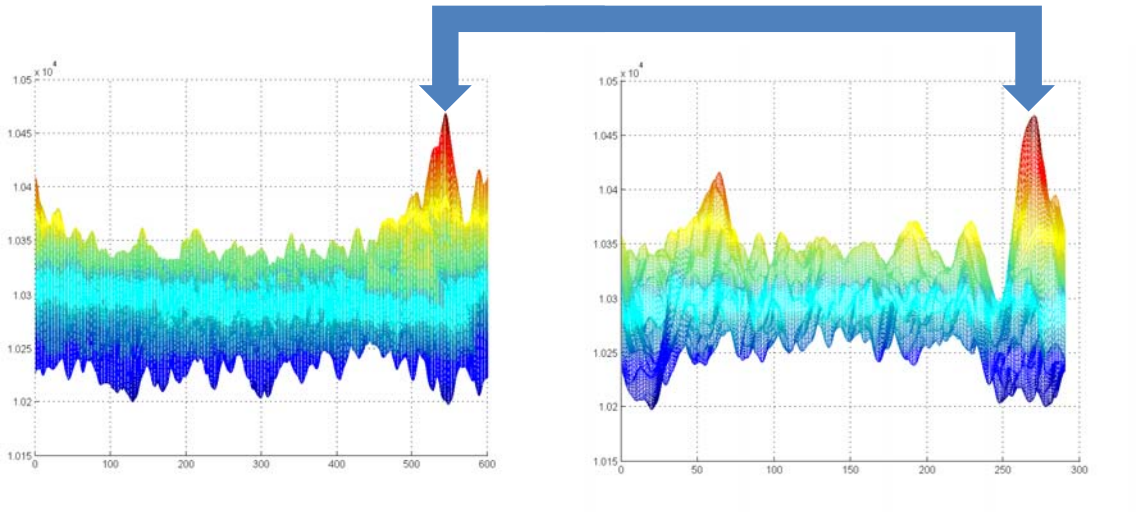
For conversion to and from the frequency domain, the origin of the image is taken at the center of the image whereas the image's reference point resides in the upper left corner, therefore the peak location of the cross-correlation matrix must be correct. The following relations are used to offset the translation results, given the peak location of the cross-correlation matrix as $(x_p, y_p)$ and the number of rows and columns in the original image as $m$, and $n$, repsectively.

$$x_t = \begin{cases} x_p - 1 & x_p < \frac{n}{2} \\ x_p - n - 1 & x_p \geq \frac{n}{2} \end{cases} \tag{5.10}$$

$$y_t = \begin{cases} y_p - 1 & y_p < \frac{m}{2} \\ y_p - m - 1 & y_p \geq \frac{m}{2} \end{cases} \tag{5.11}$$




a.



b. c.

Fig. 5.4 Cross-correlation matrix showing a peak at (543,269) which indicates a translation of (-58,-22) is required. The XYZ (a), XZ (b) and YZ(c) views are provided for illustrative purposes

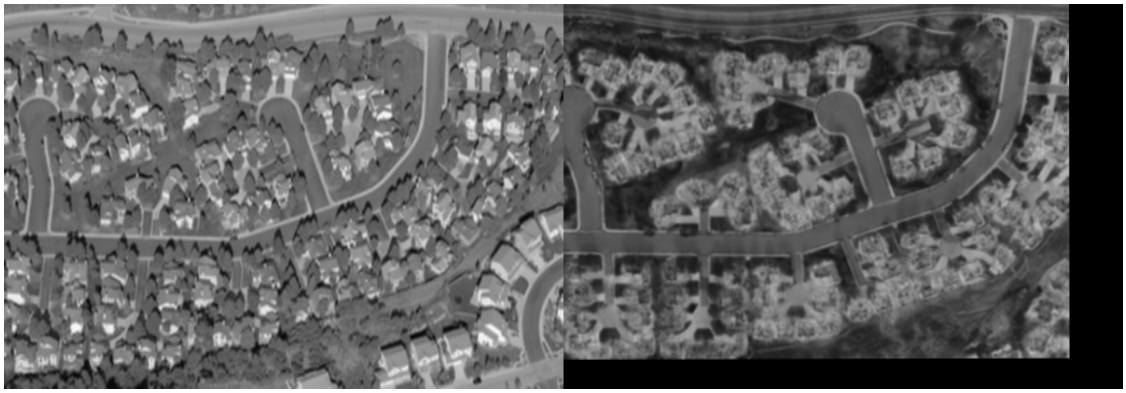


Fig. 5.5 Aerial image depicting a residential neighborhood after a fire. Original reference image (left) and translated query image (right)

An example cross-correlation matrix is visualized in Fig. 5.4 where the peak identifies the translation parameters. Using the identified parameters, the resulting coarse registration is provided in Fig. 5.5.

### 5.1.2. Rotation Estimation

The Fourier Shift theorem can be further employed to determine the rotation and scaling parameters that relate two images. To do so, the edge map for each image is used in the log-polar domain. Given two images in the spatial domain, $I_1(x, y)$ and $I_2(x_r, y_r)$ where $I_2$ varies from $I_1$ by a rotational displacement, $\theta_r$, the relationship is defined as,

$$I_2(x_r, y_r) = I_1(x\cos\theta_r - y\sin\theta_r, x\sin\theta_r + y\cos\theta_r). \quad (5.12)$$

Computing the Fourier transform of both sides yields,

$$\bar{I_2}(u, v) = \bar{I_1}(u\cos\theta_r - v\sin\theta_r, u\sin\theta_r + v\cos\theta_r). \quad (5.13)$$

After converting $\bar{I_1}$ and $\bar{I_2}$ to polar coordinates, it can be seen that the rotation in the spatial domain will be observed as a translation in the frequency domain using polar coordinates.

$$\bar{I_2}(r, \theta) = \bar{I_1}(r, \theta - \theta_r). \quad (5.14)$$

Thus, the process for recovering the rotational differences between two images involves the same procedure outlined in 5.1.1 but utilizing polar instead of spatial coordinates.

#### 5.1.2.1. Canny Edge Detection

In [73] John Canny proposes an edge detection technique that is ubiquitously known as the Canny Edge detector which meets the following three criteria.

1. Good detection that distinguishes the maximum number of edges within an image.

2. Good localization to ensure the detected edge is within a small proximity of the original edge in the image.

3. Minimal response is desired to guarantee that an edge is only detected once while neglecting to detect noise as false edges.

With the aim of reducing noise that may introduce false edges, the image is initially convolved with a Gaussian function. The discrete smoothing kernel used to estimate equation 5-2 with $\sigma_x = \sigma_y = 1.0$, is defined to be,

$$G = \frac{1}{273}\begin{bmatrix} 1 & 4 & 7 & 4 & 1 \\ 4 & 16 & 26 & 16 & 4 \\ 7 & 26 & 41 & 26 & 7 \\ 4 & 16 & 26 & 16 & 4 \\ 1 & 4 & 7 & 4 & 1 \end{bmatrix}. \tag{5.15}$$

After noise reduction, the image gradients are determined using a pair of Sobel filters. The gradient masks in the x-direction and y-direction are as follows [74],

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \tag{5.16}$$

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{5.17}$$

Gradient magnitude and direction can then be computed from the following equations,

$$G = \sqrt{G_x^2 + G_y^2} \tag{5.18}$$

$$\theta_G = \tan^{-1}\frac{G_y}{G_x} \tag{5.19}$$

A search through the gradient magnitudes, $G$, reveal edge pixels through a non-maximum suppression approach. For a given pixel with gradient magnitude, $G(x, y)$, the corresponding angle, $\theta_G(x, y)$, and neighboring pixels are considered to validate the current pixel as belonging to an edge. The following criteria are used to classify each pixel in the image where each angle is quantized to $0^\circ, 45^\circ, -135^\circ, \pm 90^\circ, -45^\circ, 135^\circ$ and the pixel $p$ belonging to the set of edge pixels is denoted as $p \in E$.

- If $\theta_G(x, y) = 0$,
  $\{p \in E \mid G(x, y) > G(x-1, y), G(x, y) > G(x+1, y)\}$
- If $\theta_G(x, y) = 45^\circ, -135^\circ$,
  $\{p \in E \mid G(x, y) > G(x-1, y-1), G(x, y) > G(x+1, y+1)\}$
- If $\theta_G(x, y) = \pm 90^\circ$,
  $\{p \in E \mid G(x, y) > G(x, y-1), G(x, y) > G(x, y+1)\}$
- If $\theta_G(x, y) = -45^\circ, 135^\circ$,
  $\{p \in E \mid G(x, y) > G(x-1, y+1), G(x, y) > G(x+1, y-1)\}$

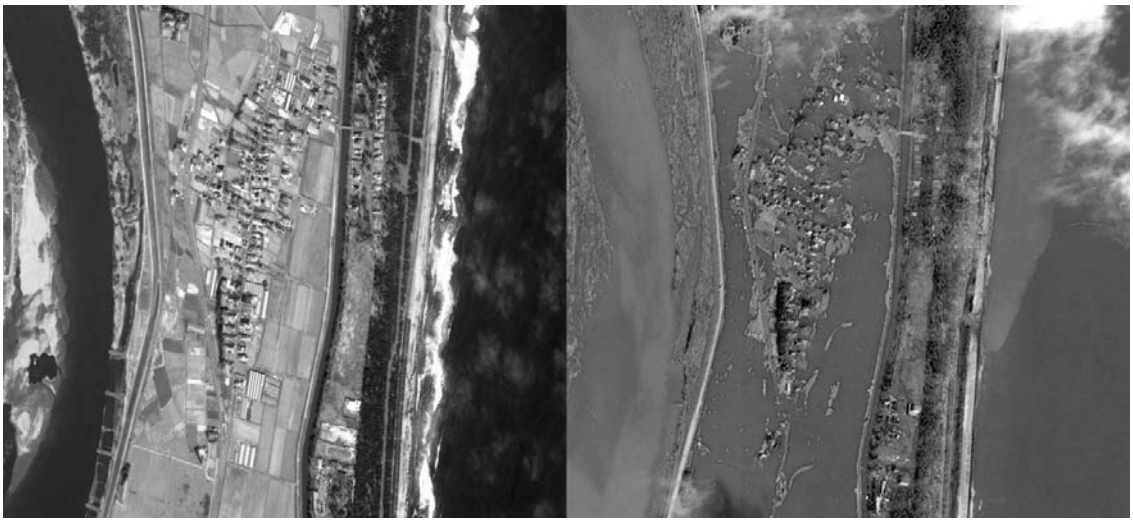
The resulting set of edge pixels, $E$, represents a binary image called thin edges. Two additional passes through $E$ is conducted where two thresholds are used to produce the final edge map. The first pass traces edge pixels while maintaining edge pixels above the high threshold and connected to an edge. It is conceivable that there will be identified edge pixels with a large gradient magnitude, however if these pixels are not connected to an edge, they are discarded. Similarly, a second pass is performed to eliminate any edge pixel below the low threshold. This process is known as edge trace hysteresis.

Edge trace hysteresis has the distinct disadvantage of using thresholds which cannot be generalized for all images; therefore the proposed registration method adopts a similar method as discussed in [75]. Let $I(x,y)$ denote a grayscale image. Then the thresholds used are calculated from the following,

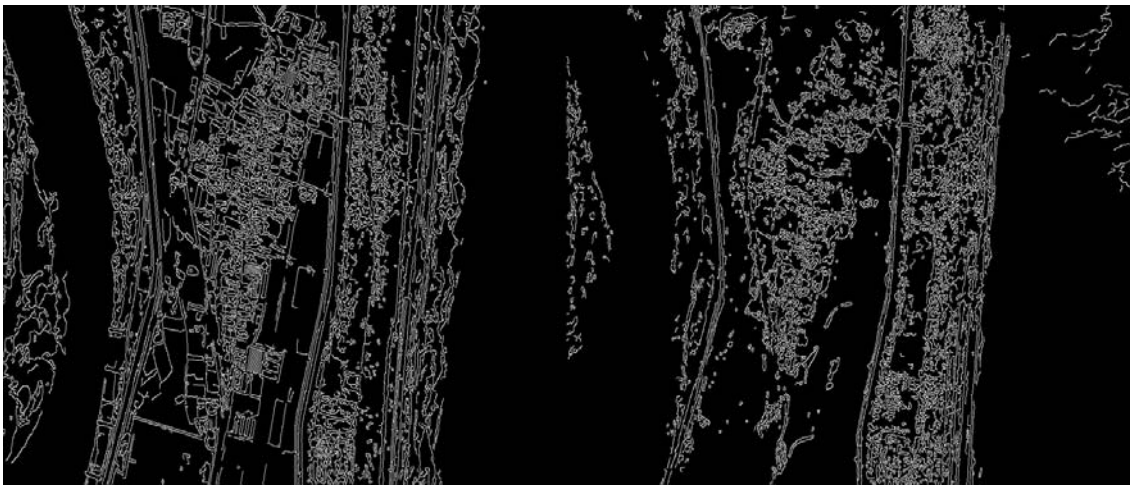
$$\tau_{low} = \frac{2}{3} * \frac{\sum_x \sum_y I(x,y)}{|I(x,y)|} \tag{5.20}$$

$$\tau_{high} = \frac{4}{3} * \frac{\sum_x \sum_y I(x,y)}{|I(x,y)|} \tag{5.21}$$

Fig. 5.6 provides an example image and its corresponding Canny edge map that was generated using the Gaussian filter in 5-15 and the Canny hysteresis thresholds (5.20) and (5.21).



a. Original reference (left) and query (right) images



b. Canny edge maps for reference (left) and query (right) images

Fig. 5.6 Example original (a) and calculated Canny edge maps (b)

#### 5.1.2.2. Morphological Operations

Mathematical morphology operators in image processing are used to describe portions of the image using shapes instead of individual pixels. The operators are applied using set theory and can all be decomposed into two basic operations, erosion and dilation [76,77]. While many applications apply morphology to binary images, their uses have since been extrapolated to grayscale and color imaging.

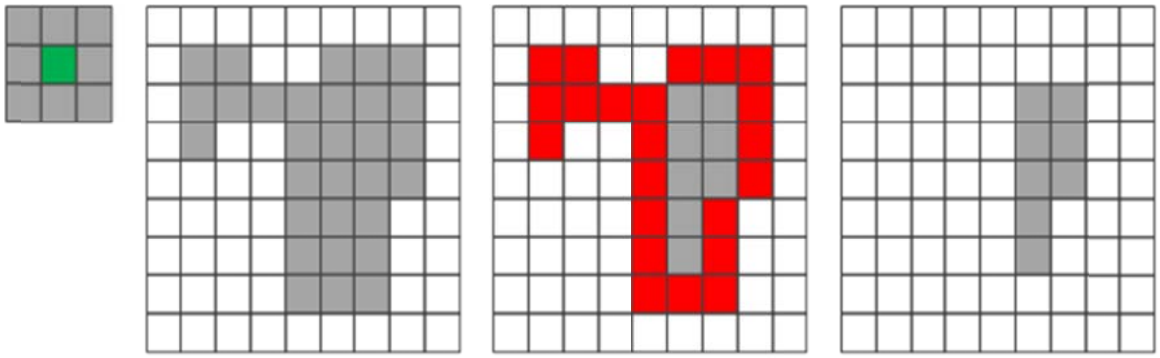
For the proposed coarse registration phase, morphological operations are applied to the image edge maps that are used for rotational estimation. They are used such that major features of an image, such as the side of a building or a major roadway, are preserved while smaller features that may negatively affect the registration outcome are removed. The aim of the operators is to reduce the image to a few large distinct features instead of many small features, as illustrated in Fig. 5.6b.

#### 5.1.2.2.1. Erosion

The erosion operator is used to remove elements from an image, using a specific structuring element, where a structuring element is a small pattern of pixels used as a template kernel. Stated formally, the pixel at $(x, y)$ is the minimum value of all pixels contained within the structuring element that is centered at $(x, y)$. If we let $S_E$ represent the structuring element, and $I$ denote the binary image, erosion can be defined as,

$$I \ominus S_E = \left\{ \min_{(x,y) \in S_E} I \mid S_E \subset I \right\} \tag{5.22}$$

To provide an example, let the structuring element be a $3x3$ block depicted in Fig. 5.7a and the binary image is given in Fig. 5.7b. The result from the erosion of $I$ by $S_E$ is shown in Fig. 5.7c where the elements removed from $I$ are identified by red elements and the center of the structuring element is denoted in green. The resulting image is provided in Fig. 5.7d.



a. b. c. d.

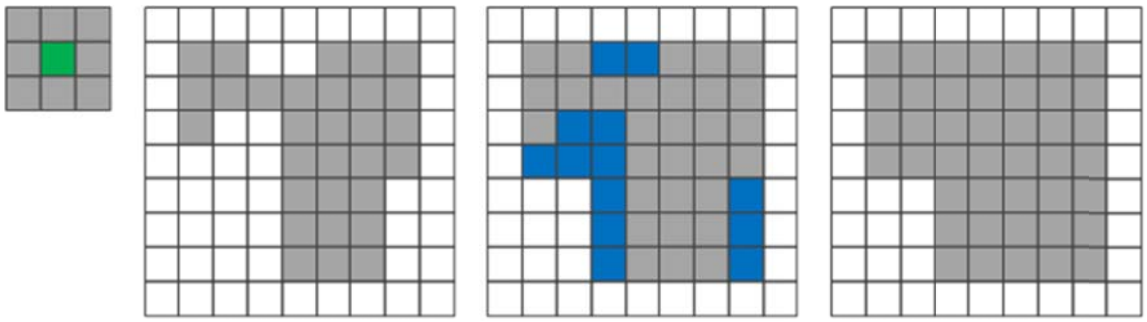
Fig. 5.7 (a) Structuring element $S_E$ used by the erosion operator, (b) binary image $I$, (c) illustration identifying the image elements removed by the operator, and (d) the resulting image after erosion by $S_E$

**5.1.2.2.2. Dilation**

The converse to erosion is dilation. This morphological operator is utilized to add elements to an image using a predefined structural pattern, $S_D$. The pixel at $(x, y)$ is the maximum value of all pixels contained within the structuring element that is centered at $(x, y)$. If we let $S_D$ represent the structuring element, and $I$ denote the binary image, dilation is defined as follows,

$$I \oplus S_D = \left\{ \max_{(x,y)\in S_D} I \mid S_D \subset I \right\} \tag{5.23}$$

Fig. 5.8 is provided as an example of the dilation operation where the structuring element, $S_D$ is a $3x3$ pattern. Elements denoted in blue indicate an element that was added to the image, and the center of the structuring element is shown in green, while the final representation is given in Fig. 5.8d.



a. b. c. d.

Fig. 5.8 (a) Structuring element $S_D$ used by the dilation operator, (b) binary image $I$, (c) illustration identifying the image elements added by the operator, and (d) the resulting image after dilation by $S_D$

It should be noted that the operator definitions in (5.12) and (5.12) state that the elements of $I$ are only affected by the structuring pattern if the structuring pattern is

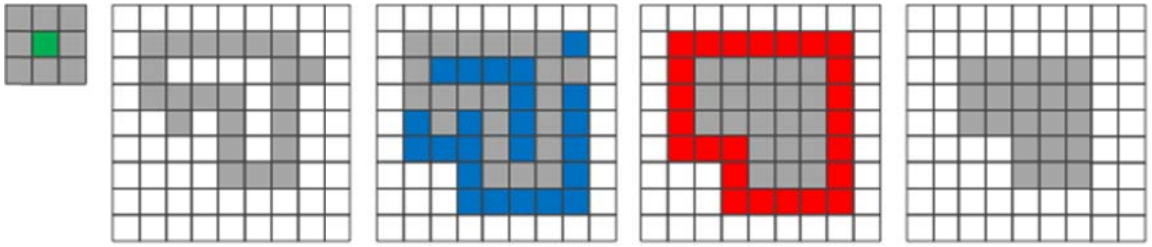
completely contained within $I$. Due to the origin of the structuring element being defined as the center, the border elements in the previous examples are not considered as the entire structuring element is not a subset of $I$.

#### 5.1.2.2.3. Morphological Close

The fundamental morphological operators, erosion and dilation, can be used to define other useful operators. Morphological close is such an operator that involves dilation followed by erosion. Formally defined, the close operation $C$ is,

$$C(I, S_D, S_E) = (I \oplus S_D) \ominus S_E. \tag{5.24}$$

The close operator can be leveraged to maintain the general shape of a group while eliminating small disjoint regions within the shape. Fig. 5.9 gives an example of the closing operation while using structuring elements $S_D$ and $S_E$, where $S_D = S_E$.



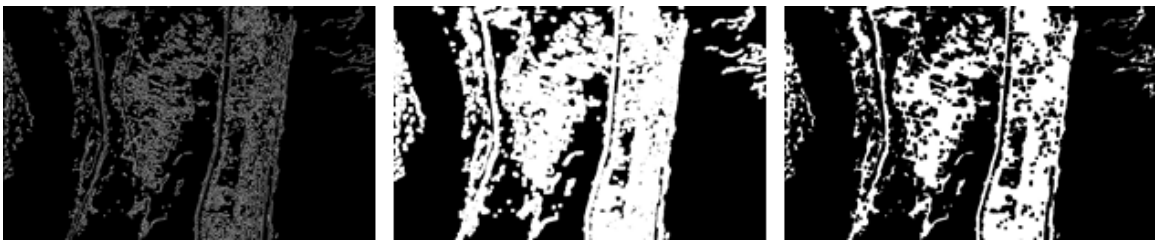
a. b. c. d. e.

Fig. 5.9 (a) Structuring elements, $S_D$ and $S_E$, (b) binary image $I$, (c) dilation of $I$, (d) erosion of dilated $I$, and (e) resulting binary image from dilation followed by erosion

As seen in Fig. 5.9e, the general shape of the original image is preserved, however the shape did become smaller in the process, therefore in the proposed system the morphological operations performed are closing followed dilation. The operation, $M$, in

equation (5.25) will preserve the main objects in an image while eliminating disjoint internal regions and external nubs. An example canny edge map is given in Fig. 5.10a along with the resulting binary image from applying the $M(I, S_D, S_E)$ operator.

$$M(I, S_D, S_E) = \big((I \oplus S_D) \ominus S_E\big) \oplus S_D. \tag{5.25}$$



a. b. c.

Fig. 5.10 (a) Original canny edge map, (b) binary image after closing operation, and (c) binary image after dilating closed image according to $M(I, S_D, S_E)$

### 5.1.2.3. Log-Polar Representation

An image in the spatial domain, $(x, y)$ is related to an image in the log-polar domain, $I(\rho, \theta)$ using the following transform,

$$\rho = \log\left(\sqrt{x^2 + y^2}\right) \tag{5.26}$$

$$\theta = \tan^{-1}\frac{y}{x} \tag{5.27}$$

Log-polar representation allows the recovery of scale and rotation parameters while exploiting frequency domain analysis. In the proposed coarse registration method, the altered canny edge map's coordinates are transformed to polar coordinates while using bilinear interpolation. The derived edge maps accurately model the shapes of the

dominant objects within a particular scene while suppressing potential details which may negatively affect the frequency domain analysis.

a. b. c.

Fig. 5.11 (a) Original image, (b) image edge map after morphological operations, and (c) log-polar edge map

## 5.2. Registration Evaluation

After the coarse registration stage and subsequent registration attempts, the registered query image is compared to the original query image in order to determine the effectiveness of that particular registration approach. The hypothesis is that an image's color histograms should be similar before and after registration. For images in the RGB color space, there are three histograms associated with the image. Each color channel's histogram consists of 256 bins while each distribution is normalized. The normalized histogram for an arbitrary color channel is given in (5.28) while an example set of histograms for a query image and correctly registered query image is given in Fig. 5.12, while Fig. 5.13 provides example histograms for a query image and an incorrect registration of the same image.

$$h_x(n) = \frac{\sum_{x,y} |I(x,y) = n|}{|I(x,y)|} \tag{5.28}$$

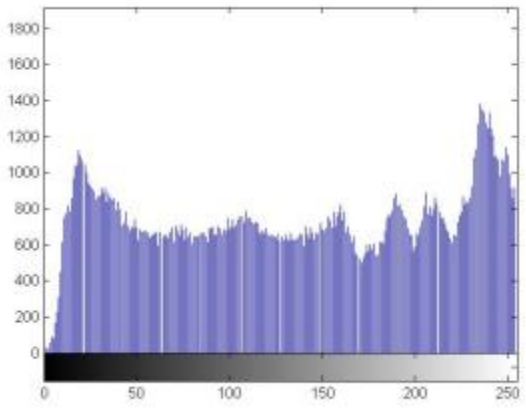
Let $h_x^1$ and $h_x^2$ represent the histograms of images 1 and 2, respectively, for color channel $x$. Then the similarity between the two images for an arbitrary 8-bit color channel, $x$, is defined as,

$$S_h^x = \sqrt{\sum_{i=0}^{255} \left(h_x^1(i) - h_x^2(i)\right)^2} \tag{5.29}$$

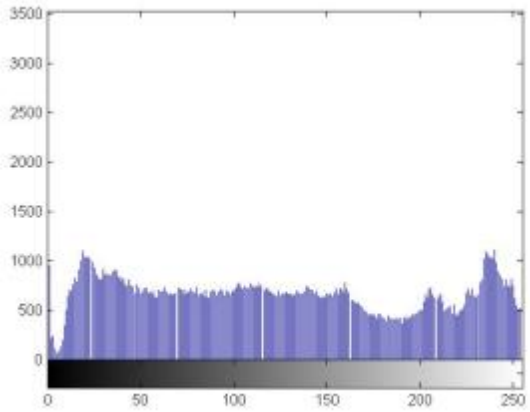
Evaluating two color images in the RGB color space requires the calculation of three similarity scores; therefore the proposed registration verification method imposes a threshold that is the sum of the three histogram scores.

$$\tau_h = S_h^R + S_h^G + S_h^B \tag{5.30}$$

Through experimentation, it has been determined that a value of 0.15 for $\tau_h$ offers a fair trade-off between registration false acceptance and rejection.




a.




d.

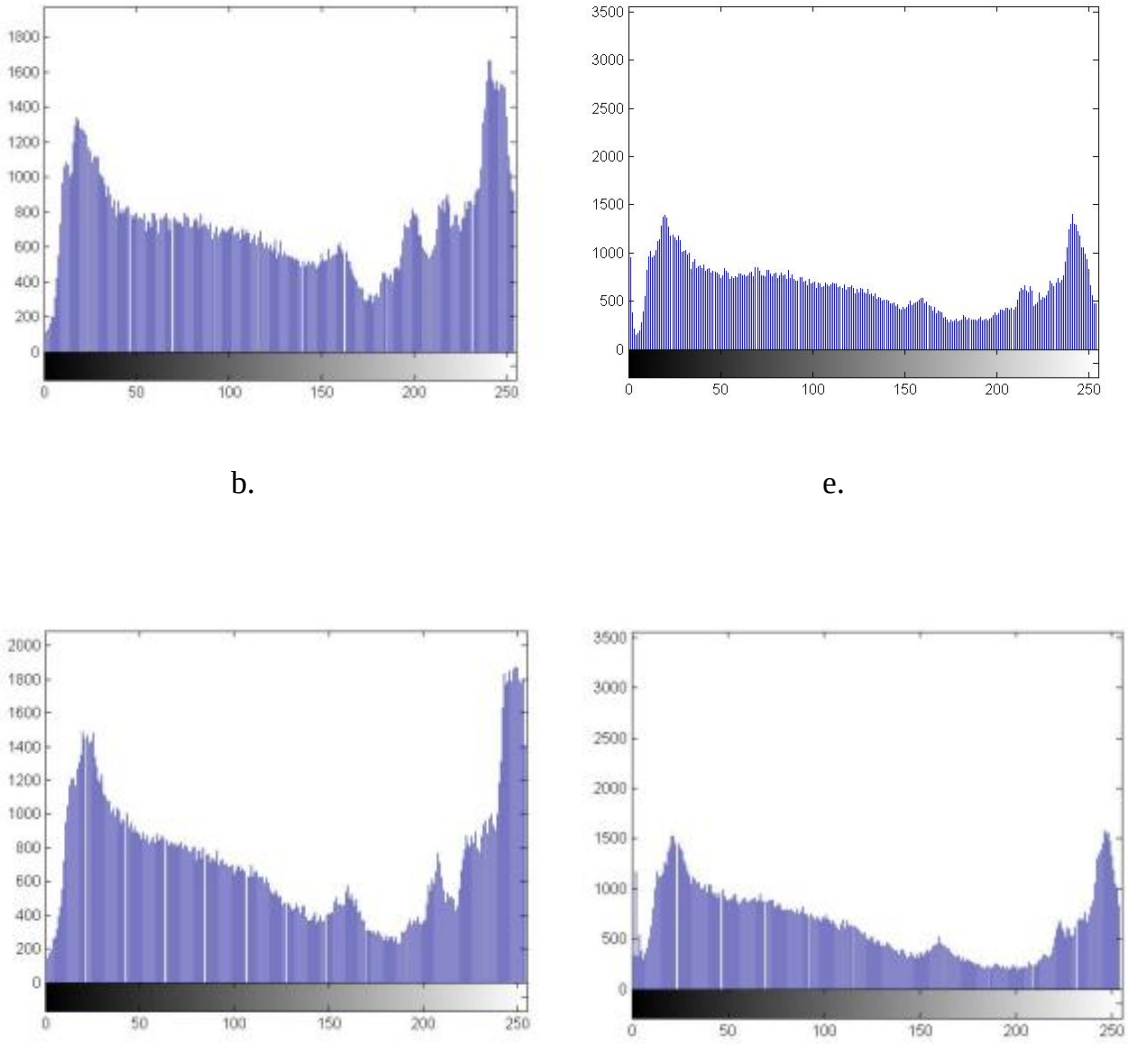



Fig. 5.12 (a-c) R,G, and B histograms for original query image, (d-f) R,G, and B histograms for accurately registered query image

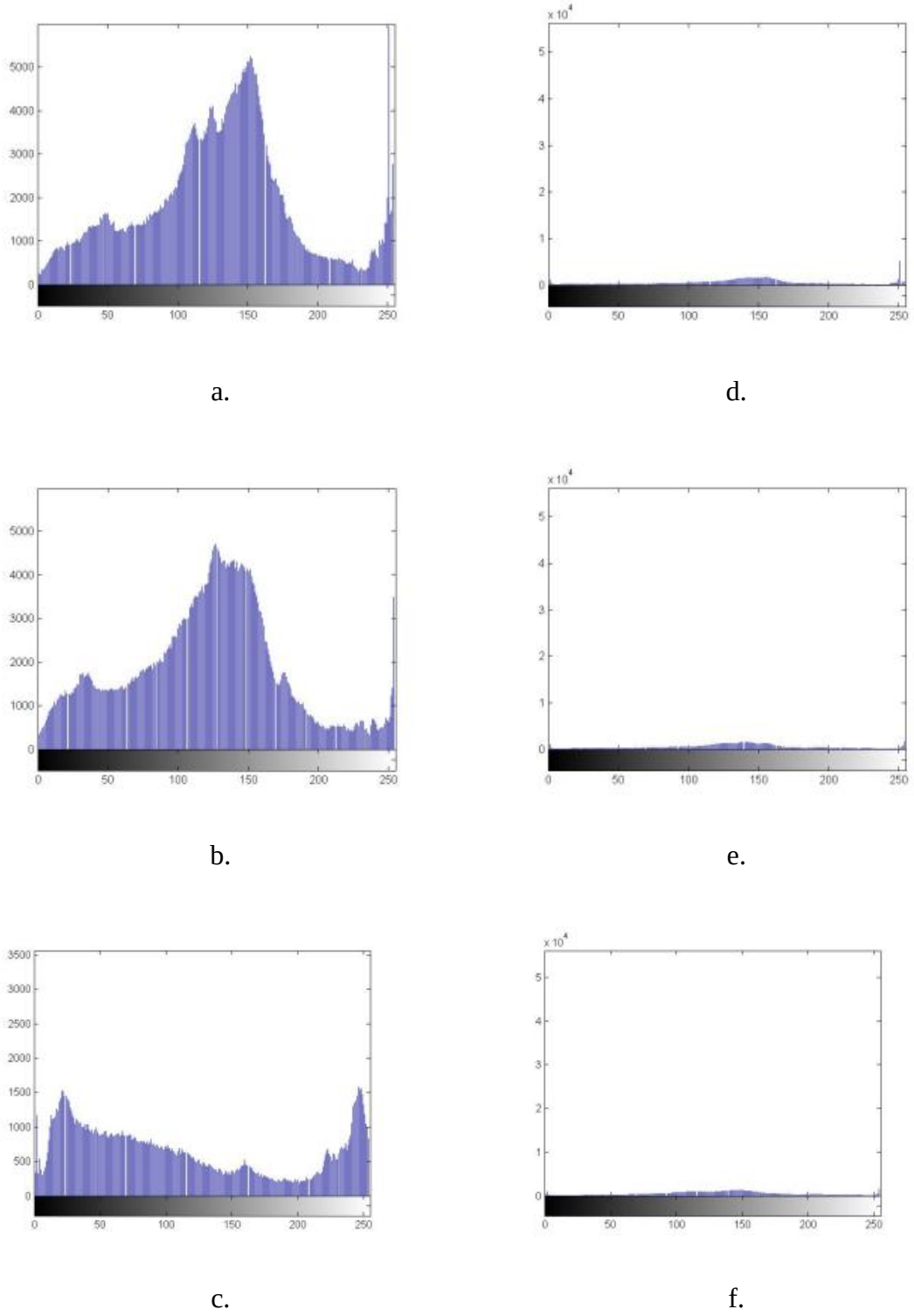



Fig. 5.13 (a-c) R,G, and B histograms for original query image. (d-f) R,G, and B histograms for inaccurately registered query image

### 5.3. Limited Neighborhood Graph-based Region Descriptor Registration

The proposed system uses an initial registration step to improve feature point matching accuracy and subsequent control point identification. If the log-polar FFT registration step is surmised to be valid, a registration attempt is performed involving the proposed graph-based region descriptor within a localized neighborhood. This limited window method is summarized in Fig. 5.14.

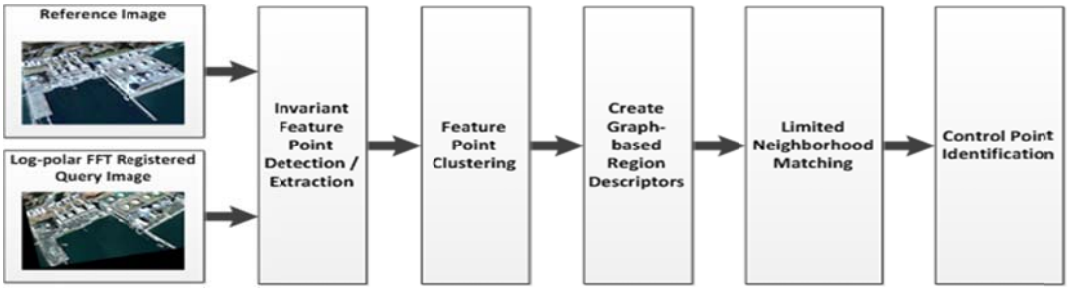



Fig. 5.14 Proposed limited neighborhood graph-based registration

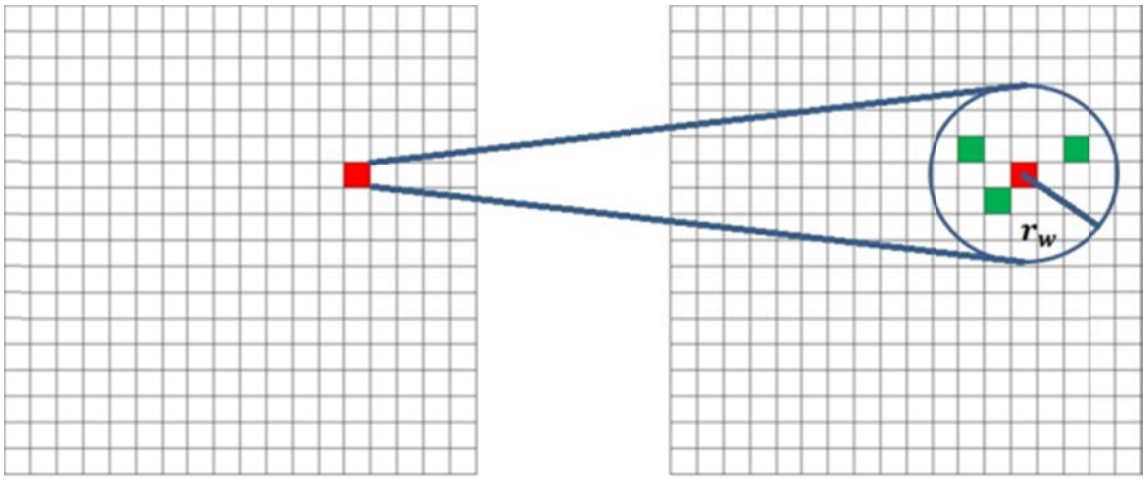
Invariant feature points are extracted from the original reference image and the log-polar FFT registered query image. For this work, k-means clustering is exploited to identify graph nodes from their spatial relationships, as outlined in Section 3.2.1.2, and the graph is constructed by using Dijkstra's shortest path algorithm that is discussed in Section 3.2.2.1. Given a graph of $n$ invariant feature points, $G = \{FP_0, FP_1, \dots, FP_{n-1}\}$, where $(x_i, y_i)$ specifies the location of $FP_i$, the centroid of the graph is defined as,

$$C_G = \left( \frac{\sum_{i=0}^{n-1} x_i}{n}, \frac{\sum_{i=0}^{n-1} y_i}{n} \right) \tag{5.31}$$

In the proposed limited neighborhood registration approach, feature points are matched utilizing a small search space of the query image that is determined from the centroid of the graph-based region descriptor in the reference image. The search radius,

$r_w$ is defined to be 20 pixels, as illustrated in Fig. 5.15. Provided a graph in the reference image with centroid at $C_r = (x_r, y_r)$, the subset of query graphs, $\overline{G_Q}$, that are tested against the reference graph is defined as,

$$\overline{G_Q} = \{G_i | \, \|C_i - C_r\| \leq r_w\}, \overline{G_Q} \subset G_Q \tag{5.32}$$




Reference Image Query Image

Fig. 5.15 Limited search neighborhood where the red square represents the centroid of the reference graph while green dots denote centroids of graphs in the query image that lie within the search window

The reference graph is then compared to all graphs in the set, $\overline{G_Q}$, where the set with smallest score is chosen as a match if the ratio between the best and second best matches is less than 0.85. For each matched pair, the centroids are acknowledged as the control points for subsequent transformation parameter estimation using the direct linear transform. Fig. 5.16 gives an example of matched features that use the log-polar FFT registration and the proposed graph-based region descriptor with a limited search window.

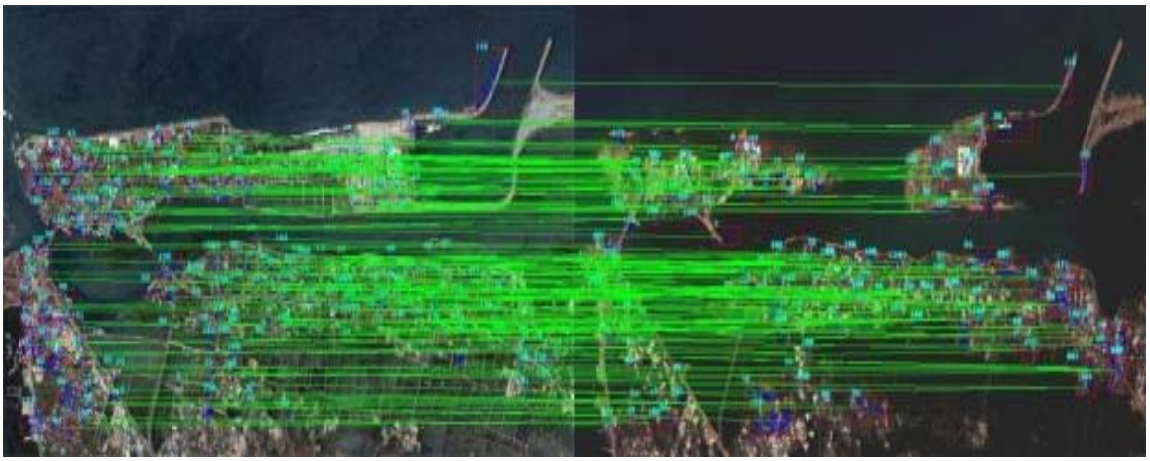


Fig. 5.16 Reference image (left) and query image (right) with matched region descriptors

### 5.4. Feature-based Registration with Coarse Area Search

The second registration approach couples intensity and feature-based approaches to accurately determine registration control points. This method is attempted if either the initial coarse registration or the limited neighborhood registration fails the histogram comparison. An overview of the approach is provided in Fig. 5.17.

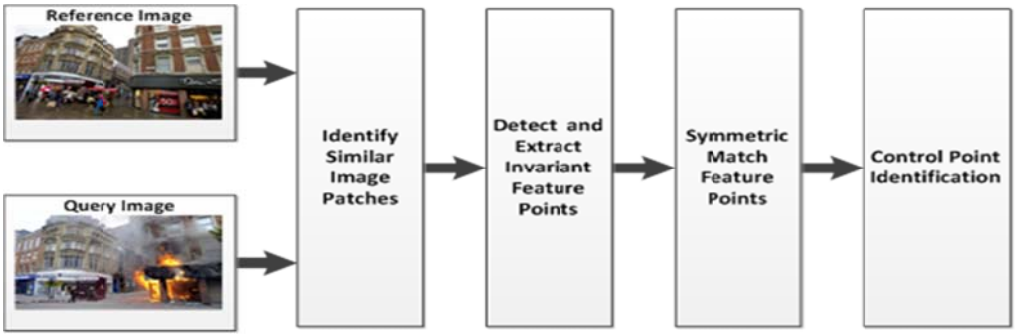



Fig. 5.17 Proposed intensity and feature-based registration method

Using the original reference and query images, the query image is segmented into $n \, x \, n$ segments. Each segment is compared to the entire reference image through a

sliding window. For each comparison, the normalized cross correlation metric is used to associate a similarity score with the two image patches. The normalized cross-correlation equation is given as,

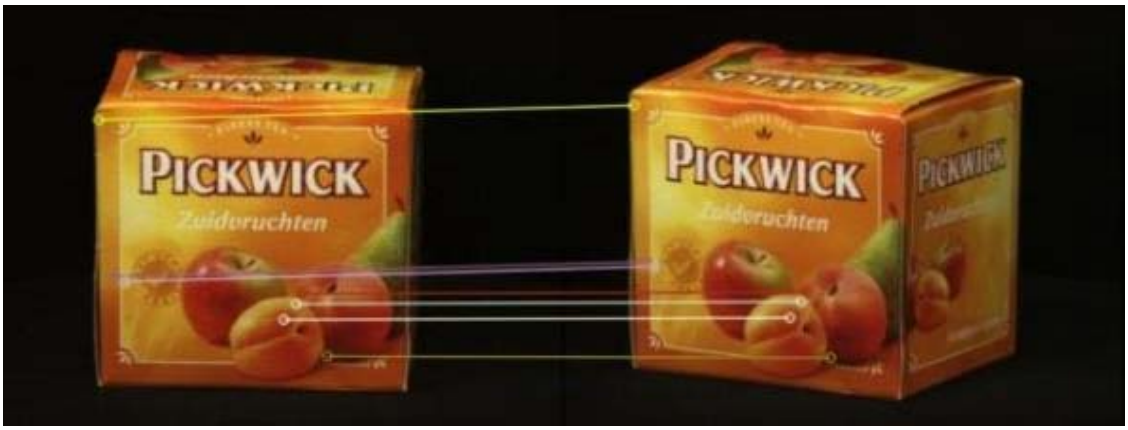
$$NCC = \frac{1}{m * n} \sum_{x,y} \frac{[I(x,y) - \bar{I}] * [t(x,y) - \bar{t}]}{\sigma_I \sigma_t}, \tag{5.33}$$

where $I(x,y)$ is the grayscale reference image, $t(x,y) \in \mathbb{R}^{m\,x\,n}$ is the image patch from the query image, whereas $\bar{I}$ and $\bar{t}$ are the means of $I(x,y)$ and $t(x,y)$, respectively. $\sigma_I$ and $\sigma_t$ denote the standard deviation of pixel intensities for $I(x,y)$ and $t(x,y)$.

For the proposed registration system, the query image is segmented such that each segment is 50 pixels wide and 50 pixels tall. Of all the matched pairs, the set of images with the 8 lowest scores are chosen for feature point matching, unless the total number of patches is less than 8 in which case all matched pairs are utilized.

Invariant feature points are then extracted and matched on a matched-pair basis, independent of any other patches. The proposed method exploits the symmetric matching scheme discussed in section 4.2 to ensure a one-to-one mapping of feature points. The set of matched feature points is then used as the set of control points for the direct linear transform. This process is shown in Fig. 5.18.

a. Matched image patches from the reference image (top row) and query image (bottom row)




b. Identified control points using SURF feature points and initial coarse template matching

Fig. 5.18 Example matched images patches and identified registration control points

## 5.5. Graph-based Region Descriptor Registration

If the comprehensive registration approach produces an inaccurate result, the final attempt to perform registration uses the proposed graph-based region descriptor without a limitation of the search space. A brute force matching scheme is executed on the original reference and query images. The entire process is summarized in Fig. 5.19 with an example set of matched graph-based region descriptors in Fig. 5.20.

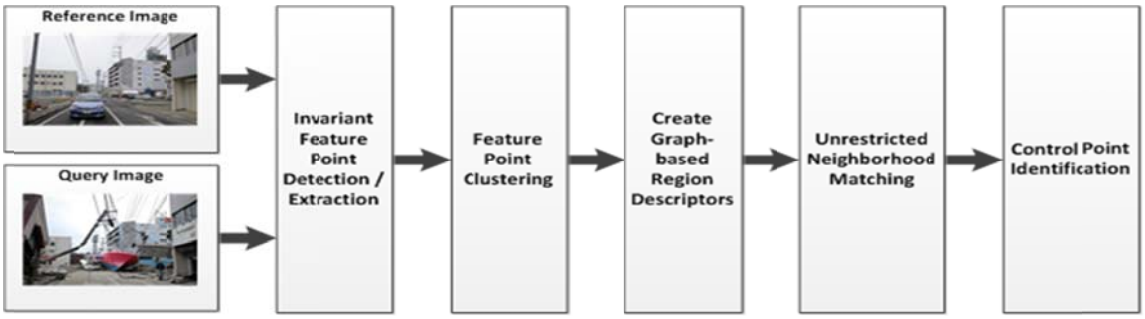



Fig. 5.19 Overview of proposed registration technique using unrestricted matching of graph-based region descriptors

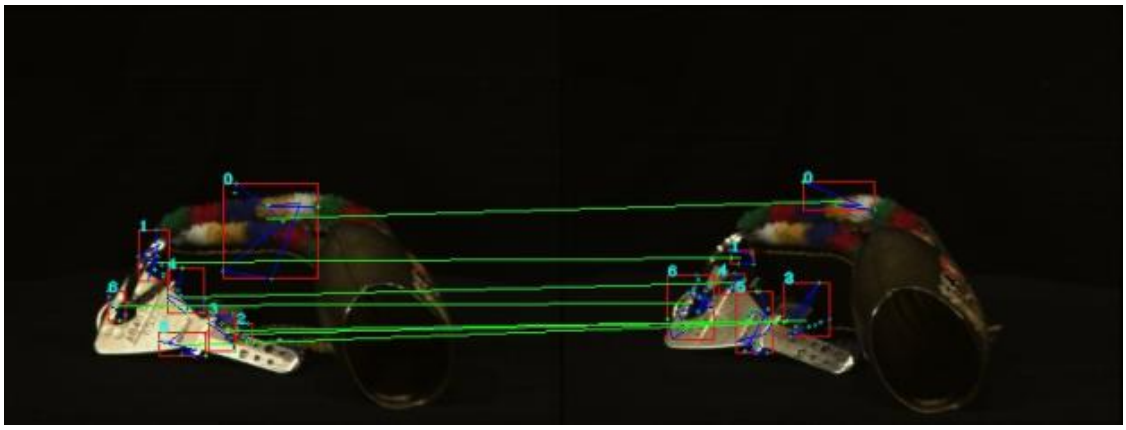


Fig. 5.20 Example matches using the unrestricted graph-based region descriptor matching method

### 5.6. Alternate Invariant Feature Points

The proposed methodologies thus far have cited invariant feature points as the basis for the graph-based region descriptor and comprehensive registration approach. More specifically, the examples and discussion outline the processes using SURF feature points; however the described framework is also capable of utilizing other invariant feature points. This section examines the Binary Robust Invariant Scalable Keypoint (BRISK) as an alternative to SURF. A detailed summary of the binary descriptor can be found in Section 2.2.1.4.

BRISK offers several advantages over SURF which suggests the descriptor should be considered for the proposed methodology. The structure of the descriptor consists of a bit-string that reduces the amount of space required to store the descriptor and greatly reduces the latency for comparing two descriptors. Moreover, the authors in [50] indicate comparable matching accuracy when compared to SURF while decreasing computation by an entire magnitude of time for some scenarios.

### 5.6.1. Limited Window Graph-based Registration with BRISK

Feature point detection can be based on any detector, however the authors in [50] propose the use of FAST which is a corner detector based on decision trees and non-maximal suppression. Clustering is performed using a Gaussian-based k-means method, as summarized in 3.2.2.1.

In order to create the graphs, a modified approach needs to be defined for identifying the initial node and for comparing nodes within a graph. While using SURF feature points, the initial node is determined as the node with the least distance to the average value of all nodes within a particular cluster. This idea is extended to BRISK by defining a descriptor that represents all descriptors within a group. The majority vote descriptor (MVD) is a 512-bit string defined as,

$$MVD = \{b_0, b_1, \dots, b_{511}\}. \tag{5.34}$$

Given a cluster of $k$ feature points that belong to the same cluster, $\{FP_0, FP_1, \dots, FP_{k-1}\}$, we denote bit $i$ of the majority vote descriptor as $MVD(i)$ and bit $j$ of feature point $n$ as $FP_n(i)$. Then bit $i$ of $MVD$ can be formulated as follows,

$$MVD(i) = \begin{cases} 1 & \left(\sum_{n=0}^{k-1} FP_n(i)\right) \geq \frac{k}{2} \\ 0 & otherwise \end{cases} \tag{5.34}$$

Descriptor comparison is accomplished using the Hamming distance that is defined as number of set bits from the bit-string result of a bit-wise XOR operation on two descriptors. Given two descriptor bit-strings, $d_i$and $d_j$, the XOR result, $r = d_i \otimes d_j$, is used to define the following Hamming distance.

$$H = \sum_{r} r_i \tag{5.35}$$

Using both the MVD for each cluster and the Hamming distance similarity metric, the graphs are constructed using the same greedy algorithm that is proposed in Section 5.3. Similarly, the matching process relies on the Hamming distance score instead of the Euclidean distance calculation. More specifically, the proposed similarity score for comparing two graphs, stated in equation (3.23), can be revised to reflect the Hamming distance.

$$S_{BRISK} = \frac{\sum_{|V|} S_{BFP}}{|V|} + S_{\theta} \tag{5.36}$$

$$S_{BFP} = \frac{H(d_i, d_j)}{512} \tag{5.37}$$

where $H(d_i, d_j)$ is the Hamming distance between descriptors $d_i$ and $d_j$, and $S_{BFP}$ is the normalized Hamming score for two descriptors.

### 5.6.2. Comprehensive Registration with BRISK

The comprehensive registration method also requires little modification to utilize the BRISK descriptor. A coarse search is still performed using the normalized cross-correlation template matching scheme. Feature point detection is achieved with the FAST corner detector. While the symmetric matching scheme is coupled with the Hamming distance metric for robust one-to-one feature point matching.

### 5.6.3. Unrestricted Graph-based Registration with BRISK

The unrestricted graph-based registration approach is altered in the same fashion as the limited neighborhood graph-based registration method. Feature point detection is accomplished through the FAST corner detector while a Gaussian-based k-means

algorithm is exploited for feature point clustering. A majority vote descriptor is calculated in order to identify the initial node for the graph creation process. The k-nearest neighbor method to determine the shortest-path is coupled with the Hamming distance to provide the basis in which the graphs are created. Lastly, an exhaustive search is performed to match graphs across images using the similarity score proposed in equation 5-36.

### 5.7. Summary

In this section a multi-stage image registration process is proposed. An initial coarse registration is attempted using the log-polar Fourier registration approach with image edge maps. The resulting registration is judged by the color histograms before and after registration. If deemed accurate, a limited search window with the proposed graph-based region descriptor is utilized. The registration results are again verified through color histograms. If the limited window approach fails, a comprehensive method, which couples an intensity-based coarse search and feature-based fine search for suitable registration control points, is pursued. Finally, if the comprehensive method is determined to be inaccurate, a broad search exclusively using the proposed graph-based region descriptor is conducted. Lastly, a thoroughly analysis is presented on an alternative invariant feature point which forms the basis of the proposed methods.

# 6. RESULTS AND DISCUSSION

In the following sections, the results from each significant phases of the method outlined in Chapter 5 are presented and discussed. Furthermore, the outcome of the overall proposed approach is provided for a test databases that consists of 41 image pairs of scenes before and after a disaster. The database is composed of aerial images and images depicting urban scenes. Additionally, a second database consisting of additional buildings for varying viewpoints as well as a set of general objects from the Amsterdam Object Library [72], is used to prove viability of the proposed methods to other applications. Lastly, performance benefits realized from a GPU implementation is discussed.

## 6.1. Initial Coarse Registration

The initial phase correlation registration task attempts to recover translation and rotation parameters from the cross power spectrum of two images. In order to analyze the effectiveness of the transformation parameter estimation, the reference image from each image pair is rotated between $1^{\circ}$ and $180^{\circ}$, in $1^{\circ}$ increments. The rotated reference image is then registered to the original reference image. The root mean square error (RMSE) is calculated for each image set according to,

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}\left(x_e^i - x_m^i\right)^2}, \tag{6.1}$$

where $n$ is the number of iterations (180), $x_e^i$ is the $i$-th expected value, and $x_m^i$ is the $i$-th measured value.

In the proposed approach, the rotation parameter is estimated using the Canny edge maps for a given image pair. The edge maps are shown to be more accurate when determining the rotation between two images as shown in the comparison of Fig. 6.1. The graphs depict the RMSE versus angle of rotation for both the edge map comparison and the original image comparison. For each angle in the range $1^\circ$ to $180^\circ$, the RMSE was calculated for each of the 41 image pairs. The average RMSE using the Canny edge maps and average RMSE using the original images, is provided below. It is clearly realized that the edge map is superior for rotation estimation.

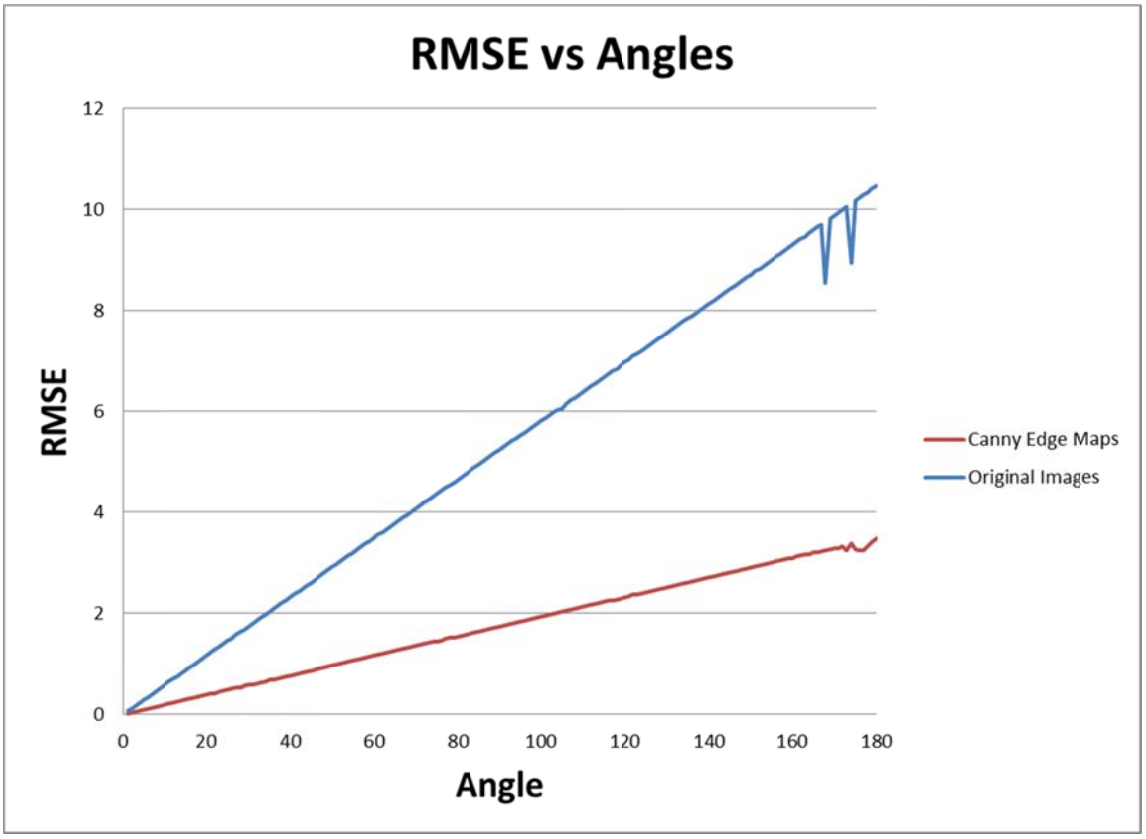



Fig. 6.1 RMSE for rotation angles ranging from $1^\circ$ to $180^\circ$

Similarly, a set of controlled translations is applied to the original reference image without any rotational alteration to produce a ground truth. The translations exploited are all combinations of 1-50 pixels in the x-direction and 1-50 pixels in the y-direction. The RMSE for x and y translations is visualized in Fig. 6.2. It is apparent that the use of Canny edge maps do not offer any advantages over using the original images directly to recover the translation estimation, therefore the proposed approach first estimates the rotation from the down sampled edge maps then the translation parameters are calculated from the down sampled original images. Fig. 6.2 was generated by calculating the average RMSE for all 2500 combinations for each of the 41 image sets.

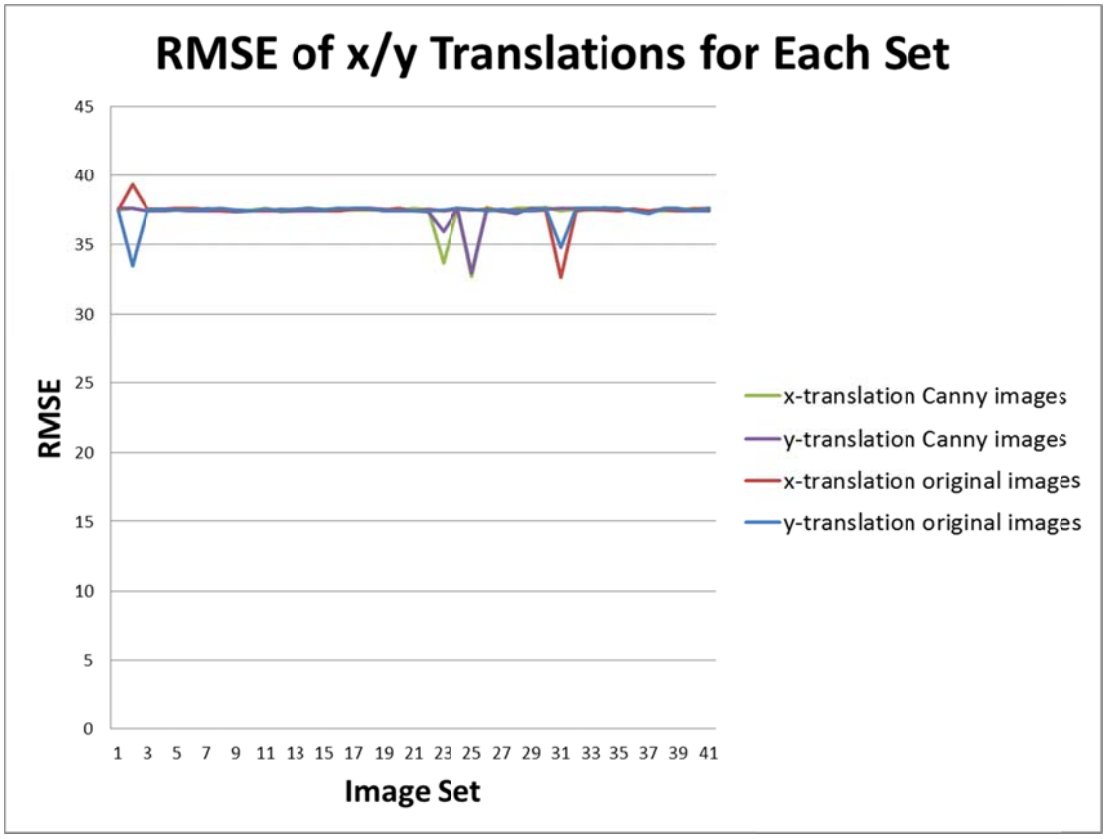



Fig. 6.2 Average RMSE for all combinations of x and y translation from 1 to 50 pixels

The coarse registration results are provided in Fig. 6.3 through Fig. 6.5, where Fig. 6.3 and Fig. 6.4 depicts an accurate registration while Fig. 6.5 provides an example of a

poor coarse registration. The poor coarse registration can be attributed to large variations in image content between the reference and query image. In the provided example, a large fire is present in the query image. This obstruction negatively affects the rotational estimation. In an ideal situation, the coarse registration is most advantageous if only geometric variations are present, whereas photometric variations such as obstructions or large amounts of noise adversely affect the results. Fig. 6.3 and Fig. 6.4 show images that vary in lighting conditions; however, the edge maps are similar enough that the translation and rotation parameters are easily recovered.

a. Original reference image

b. Original query image

c. Registered query image after phase correlation with Canny edge maps

Fig. 6.3 Coarsely registered images from an aerial view of a natural scene

a. Original reference image

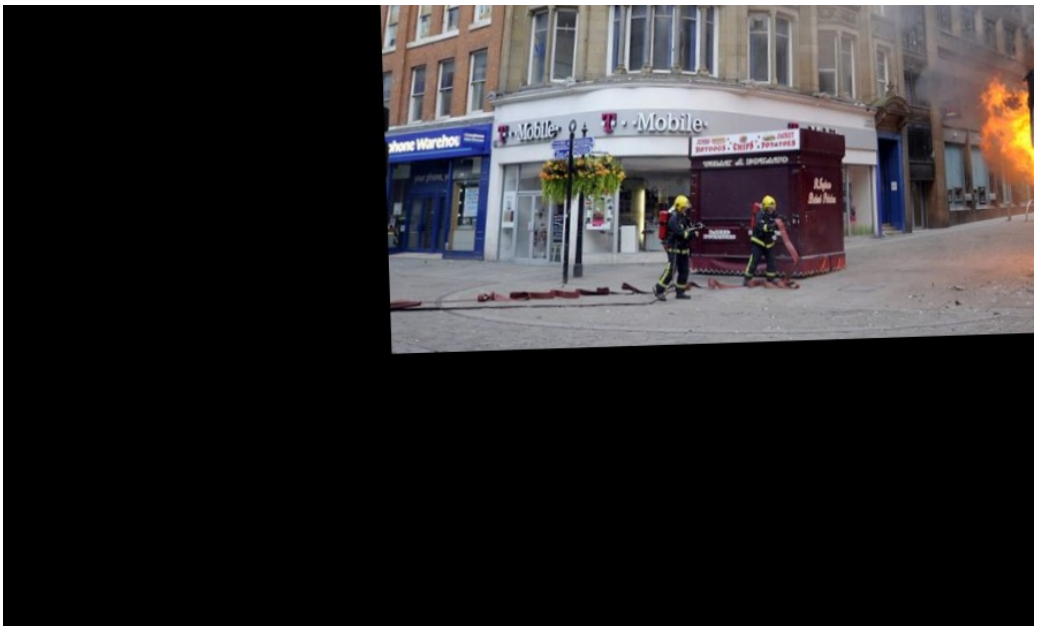
b. Original query image

c. Registered query image after phase correlation with Canny edge maps

Fig. 6.4 Coarsely registered images from an aerial view of structural content


a. Original reference image


b. Original query image



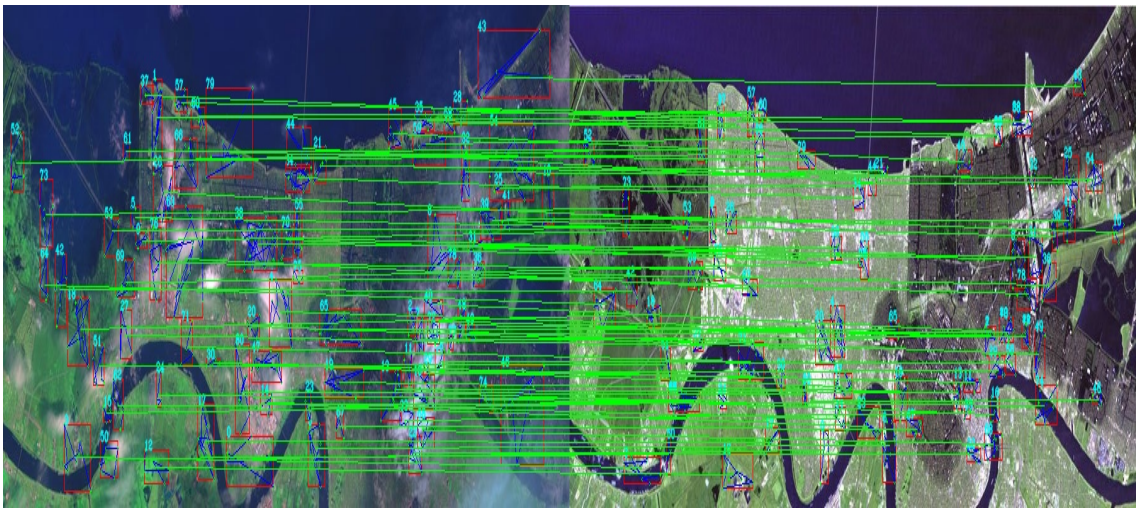

c. Coarsely registered query image.

Fig. 6.5 Failed coarse registration of an urban scene

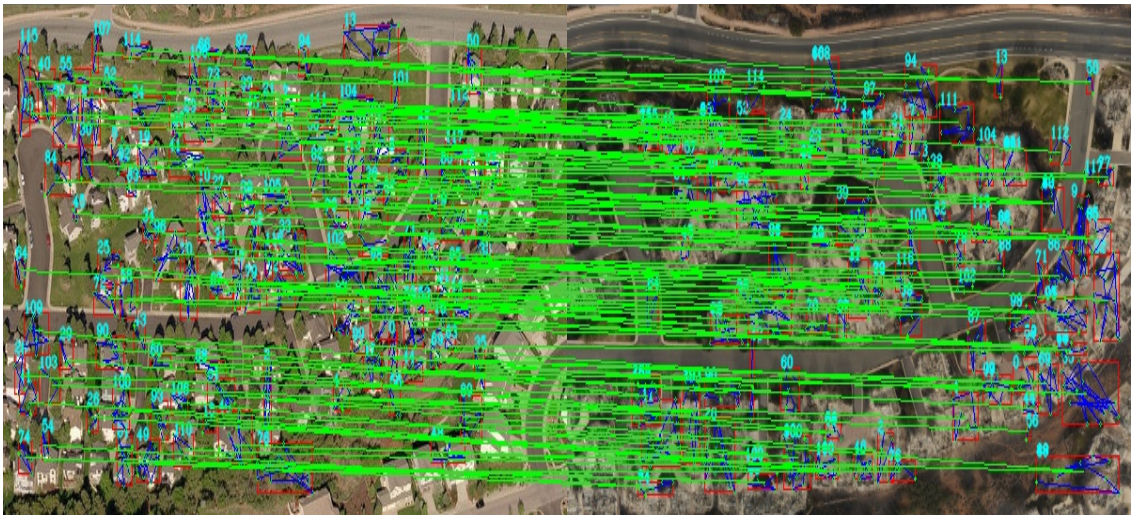
### 6.2. Limited Window Graph-based Region Descriptor Registration

The limited window graph-based region descriptor registration approach exploits the results of the log-polar phase correlation registration. If the coarse registration step is validated using the color histograms, as discussed in Section 5.2, invariant feature point clusters are used as a basis for forming k nearest neighbor (k-NN) shortest distance graphs, where $k = 1$. The registration technique's success is directly related to the matching of the graph-based region descriptors. Since the coarse registration is verified to be accurate, the search space for matching graphs is greatly reduced. Fig. 6.6 illustrates several examples of matched region descriptors.

The examples shown in Fig. 6.6 provide evidence supporting the broad applications in which the proposed region descriptor can offer accurate feature matching. Each example shows the reference image on the left while the query image is provided on the right. The detected graph-based region features are identified with red bounding boxes while the green lines illustrate the matched pairs across images. Fig. 6.6a gives an example of an aerial image with partial occlusion of a natural scene, while Fig. 6.6b and Fig. 6.6c are examples of an aerial view of residential building and of an urban scene, respectively. The region descriptor is shown to provide accurate matches for all three scenarios when both geometric and photometric differences exist between images. Furthermore, visual inspection indicates high matching accuracy as observed by the parallel relationship between green lines.



a. Region descriptor matching on an aerial image with photometric variation



b. Region descriptor matching on an aerial image with geometric variation


c. Region descriptor matching on an urban scene with geometric variation and obtrusion

Fig. 6.6 Example region descriptor matching results

For the given test database, 65.8% of the image sets rely on the limited window region descriptor registration technique. Several examples of successful registration utilizing the initial coarse registration and limited window region descriptor approach are given in Fig. 6.7 and Fig. 6.8. The first example is an aerial view of commercial structures that include content differences due to obstruction and lighting variations, while the second example is another aerial view of a residential area. Both examples are successfully registered using the limited window region descriptor method.

a. Original reference image

b. Original query image

c. Registered query image using limited search window technique

Fig. 6.7 An aerial view of structural images with large differences in pixel intensities

a. Original reference image

b. Original query image

c. Registered query image using limited search window technique, depicting an aerial view of residential structures

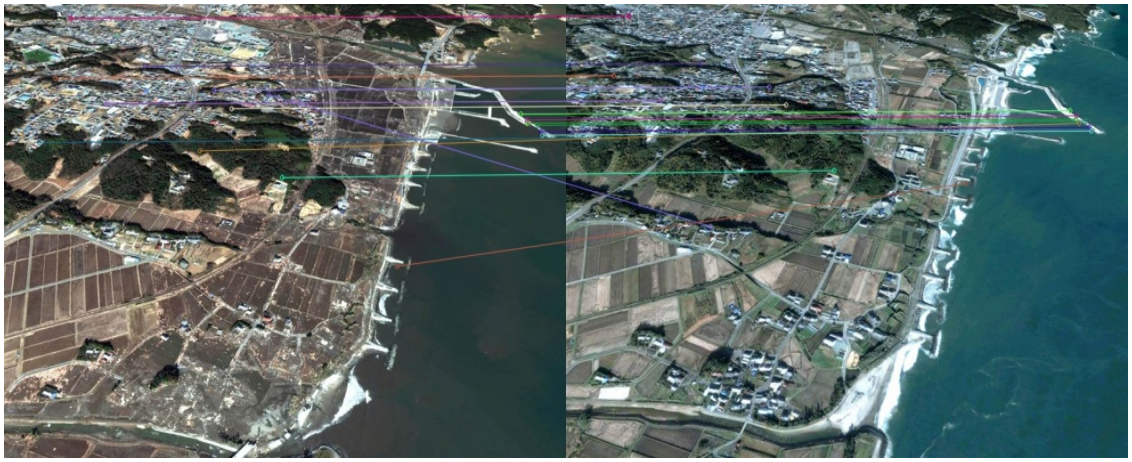
Fig. 6.8 Residential registration example using the limited window region descriptor method

### 6.3. SURF-based Registration with Coarse Template Matching

If the initial coarse registration or limited window registration approach produces invalid results, as determined by the histogram comparisons, a coarse template search is conducted using the normalized cross correlation metric to determine similar image regions. Each pair of matched regions is then matched locally using invariant feature points. For the proposed method, SURF feature points provide the basis for control point identification where a symmetric matching scheme is exploited. The following figure illustrates two examples of image sets that utilized the SURF matching with coarse initial template matching. Of the entire test database, 12.2% of the image sets were registered using the SURF feature points directly after identifying similar image patches through the use of the normalized cross-correlation metric. Fig. 6.9 shows a successful registration of

an aerial view of natural content. Pixel intensities vary slightly while the image content is from a different viewpoint. Fig. 6.10 demonstrates the effectiveness of the proposed method with a street view perspective of a building that has sustained significant damage. Although the textural properties are similar, large geometric differences are present due to the devastating effects of an earthquake. In both examples of Fig. 6.9 and Fig. 6.10, the original reference image is shown, along with the SURF matches and registered query image. The image depicting the feature point matches contains the query image on the left and reference image on the right where the matched points are connected through colored lines.

a. Original reference image



b. Matched SURF feature points

c. Registered query image using SURF feature points

Fig. 6.9 Aerial image registered using NCC template matching and SURF


a. Original reference image

b. Matched SURF feature points


c. Registered query image

Fig. 6.10 Street view of a registered building using NCC template matching and SURF

### 6.4. Unrestricted Graph-based Region Descriptor Registration

In the last mode of the proposed method, an unrestricted search is performed to match the proposed graph-based region descriptor. The resulting matching pairs are used as the basis for control point identification. This approach is only attempted if the previous two methods are determined to yield incorrect registrations. The remaining 22% of the test sets utilized the unrestricted technique. Registration results are shown in the following, Fig. Fig. 6.11 through Fig. 6.14.

Fig. 6.11 provides an example image set that is very similar geometrically yet varies greatly in pixel intensity. Additionally, the features within Fig. 6.11a are not as clearly

defined due to the poor quality of the image. This example demonstrates the value of the proposed approach in the scenario in which the two images vary in image quality.

In Fig. 6.12, the unlimited window search method is utilized to successfully register street view images of a building before and during a fire. The fire and subsequent smoke provide significant obstruction in the query image. Moreover, the images were captured at different times and therefore from different perspectives. This example exhibits the proposed approach's robustness for street view scenes during a disaster.

The example provided in Fig. 6.13 depicts the effectiveness of the proposed approach as applied to an aerial view of a residential area after a devastating fire. Lighting conditions differences are minimal; however image content differs immensely due to the fire damaged houses and overall scale of the image. This instance shows how the proposed method can be successful for images of different scales.

Lastly, the example shown in Fig. 6.14 is another street view building where the structure of the building in the query image contrasts from the original image as a result of the damage sustained from a riot. Approximately 50% of the query image differs from the reference, however the proposed region descriptor successfully identified suitable control points for registration.

a. Original reference image

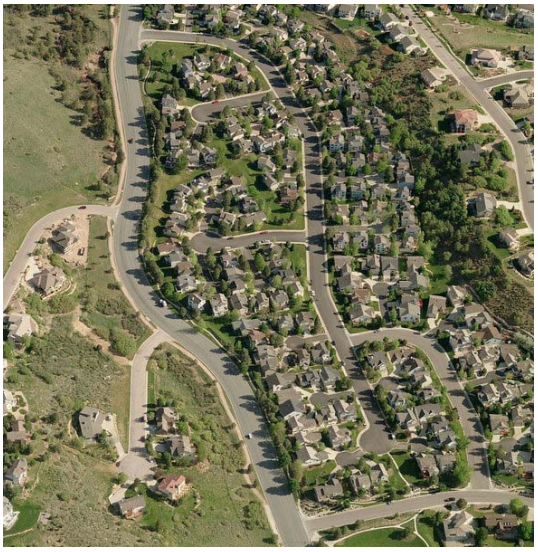
b. Registered query image

Fig. 6.11 Registration of an aerial view of a natural scene


a. Original reference image


b. Registered query image

Fig. 6.12 Registration of a street view building during a disaster



a. Original reference image

b. Registered query image

Fig. 6.13 Registration of an aerial view of a residential area greatly affected by fire damage


a. Original reference image


b. Registered query image

Fig. 6.14 Registration of a street view building after significant damage has been sustained during a riot

### 6.5. Overall Results

As another baseline for comparison, the mutual information metric was used to determine the effectiveness of the proposed algorithm as demonstrated in [66]. The function is optimally 1.0 when two images are identical and therefore must be maximized for accurate registration. For comparison, the reference image and query image are converted to grayscale while the MI score is calculated from (6.2).

$$MI(I_1, I_2) = H(I_1) + H(I_2) - H(I_1, I_2), \tag{6.2}$$

where $H(I_x)$ is the entropy measure for image x and $H(I_x, I_y)$ is the joint entropy of images x and y. $H(I_x)$ and $H(I_x, I_y)$ are defined as follows,

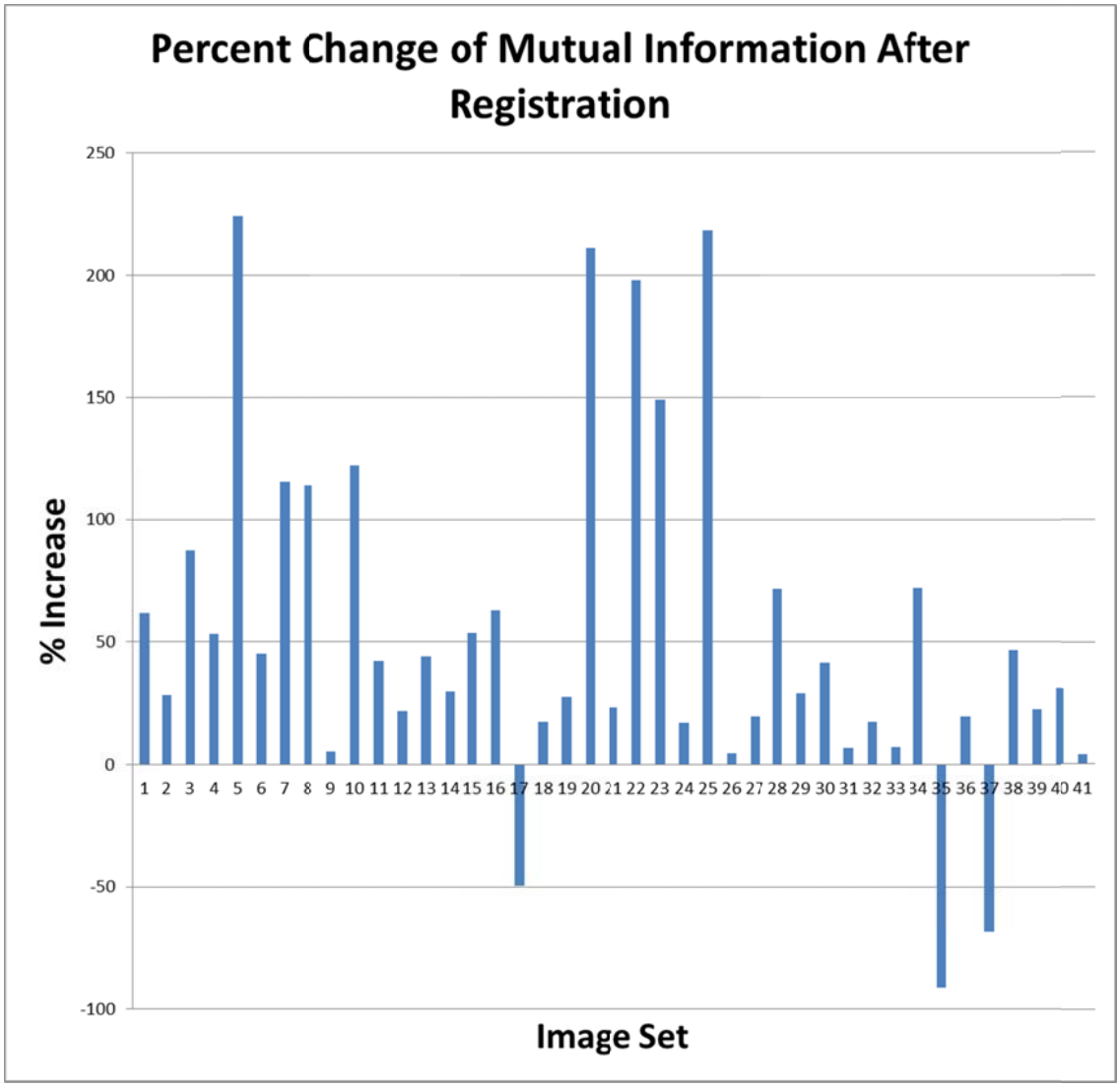
$$H(I_x) = -\sum_{x \in I_x} p_{I_x}(x) \log\left(p_{I_x}(x)\right), \tag{6.3}$$

$$H(I_x, I_y) = -\sum_{x \in I_x} \sum_{y \in I_y} p_{I_x I_y}(x, y) \log\left(p_{I_x I_y}(x, y)\right). \tag{6.4}$$

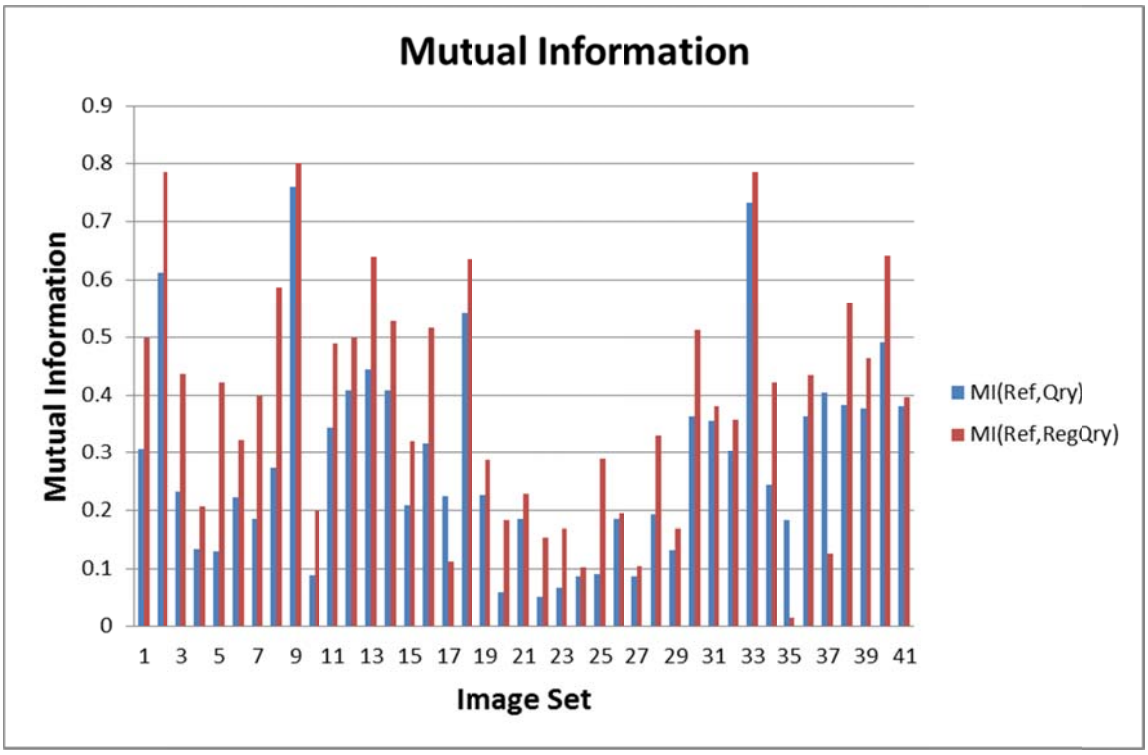
The probability density function, $p_{I_x}(x)$, is estimated from the intensity histogram of image $I_x$ while $p_{I_x I_y}(x, y)$ is calculated from the joint histogram of images $I_x$ and $I_y$.

For each set of matched image pairs, the mutual information score was calculated for the original reference image and the original query image. These scores are denoted in blue in Fig. 6.15. Similarly, the mutual information score was calculated for the reference image and the registered query image and is indicated in red. As illustrated, 92.7% of the query images produced a larger mutual information score after applying the proposed registration technique. The bar graph of Fig. 6.15a illustrates the percent change in mutual information when comparing the metric before and after registration. This visual representation confirms that the registered images improved the score for

scenarios including aerial views of commercial and residential buildings, as well as buildings from a street view. In each scenario the image content varies geometrically from perspective differences and photometric variations are present through differences in lighting conditions, obstructions and natural disasters. In the most challenging cases, successful registration is realized when image content varies greatly due to the devastating effects of a disaster, such as fire and flooding damage. From the evaluated image sets, it is shown that the proposed registration method is an effective approach for applications involving images before and after a disaster.




a. Percent increase in mutual information after the proposed registration method




b. Mutual information score for comparing reference and query images (blue), as well as reference and registered query images (red)

Fig. 6.15 Mutual information comparisons before and after registration

The test set was segmented into three classes where class 1 represent the set of images that are ideally registered using the limited window method, while class 2 and class 3 are image sets accurately registered using the comprehensive method and unlimited search window technique, respectively. The mutual information scores for each class using each approach is summarized in Table 6.1. As shown, the mutual information score is maximized for a particular class using the proposed method. Moreover, when a different approach is attempted, the mutual information is shown to be less than the score generated from the ideal approach. Table 6.1 indicates the proposed registration

methodology is effective for a wider range of applications than if a single approach is used. Moreover, the overall mutual information score using the registered query image from any of the proposed methods is observed to be greater than the score without registration in 92.7% of the tested cases. It is for the class 1 and class 2 test sets, a higher mutual information score observed using the unlimited search approach, however this is coupled with significant latencies associated with an exhaustive search.

A summary of the mutual information distribution over the entire test set is provided in Table 6.2 where the average mutual information scores for a common intensity-based registration approach and feature-based registration method are presented for comparison. The intensity-based approach employs the least-squares similarity metric while the feature-based registration method detects and matches SURF keypoints which are used with RANSAC to identify registration control points.

Table 6.2 offers strong evidence of the proposed region descriptor's effectiveness for identifying suitable registration control points. The table shows that the proposed method increased the mutual information to an average score of 0.46 for the test set, which is an increase of 61.24% over the score using the original reference and query images. Using the common registration approaches as a baseline, the proposed registration approach outperforms the SURF-based registration by 155.5% and the least-squares method by 12.2%.

Table 6.1 Mutual information scores per class compared to traditional intensity and feature-based approaches

| Mutual Information Per Classes | | | | | | |
|---|---|---|---|---|---|---|
| | No Registration | Intensity-based Registration | Feature-based Registration | Limited Window Registration | Comprehensive Registration | Unlimited Search Registration |
| **Class 1** | 0.301 | 0.480 | 0.178 | **0.512** | 0.401 | 0.508 |
| **Class 2** | 0.275 | 0.415 | 0.208 | 0.01 | **0.440** | 0.217 |
| **Class 3** | 0.284 | 0.343 | 0.156 | 0.102 | 0.078 | **0.428** |

Table 6.2 Overall average mutual information or proposed registration method compared to no registration as well as intensity and feature-based registration approaches for entire the test set

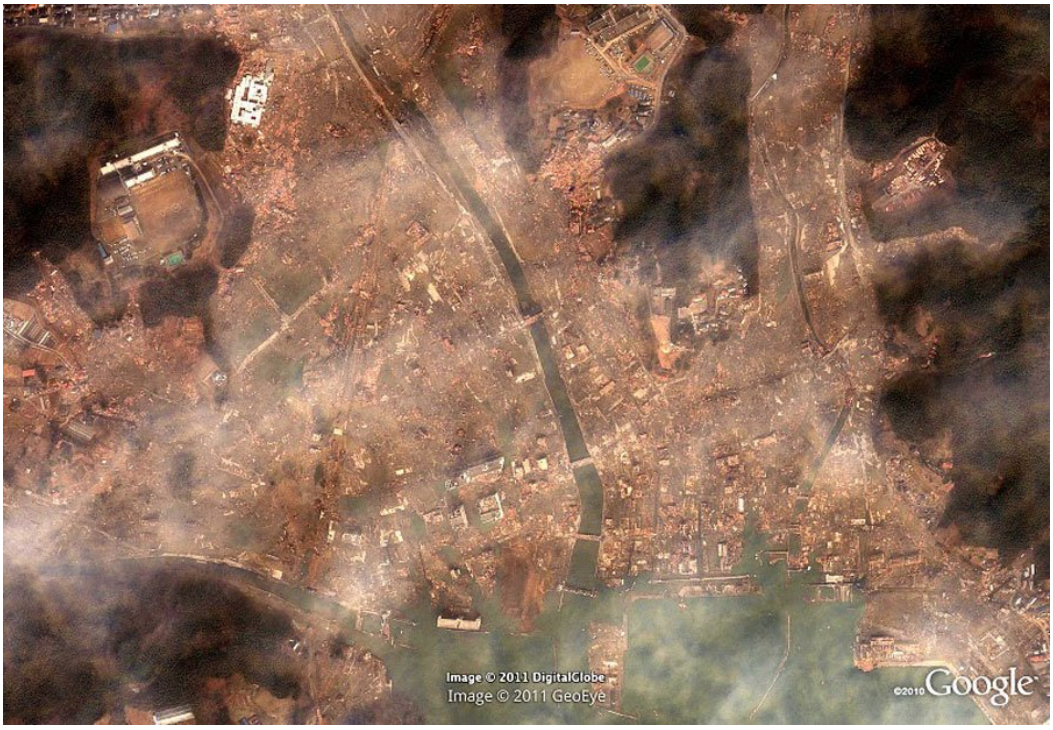
| Mutual Information Comparisons | | | | |
|---|---|---|---|---|
| | I(Ref;Qry) | I(Ref;Intensity) | I(Ref;Feature) | I(Ref;Proposed) |
| Average | 0.29 | 0.41 | 0.18 | 0.46 |
| Std Deviation | 0.17 | 0.20 | 0.21 | 0.16 |

### 6.5.1. Class 1 Result Comparisons

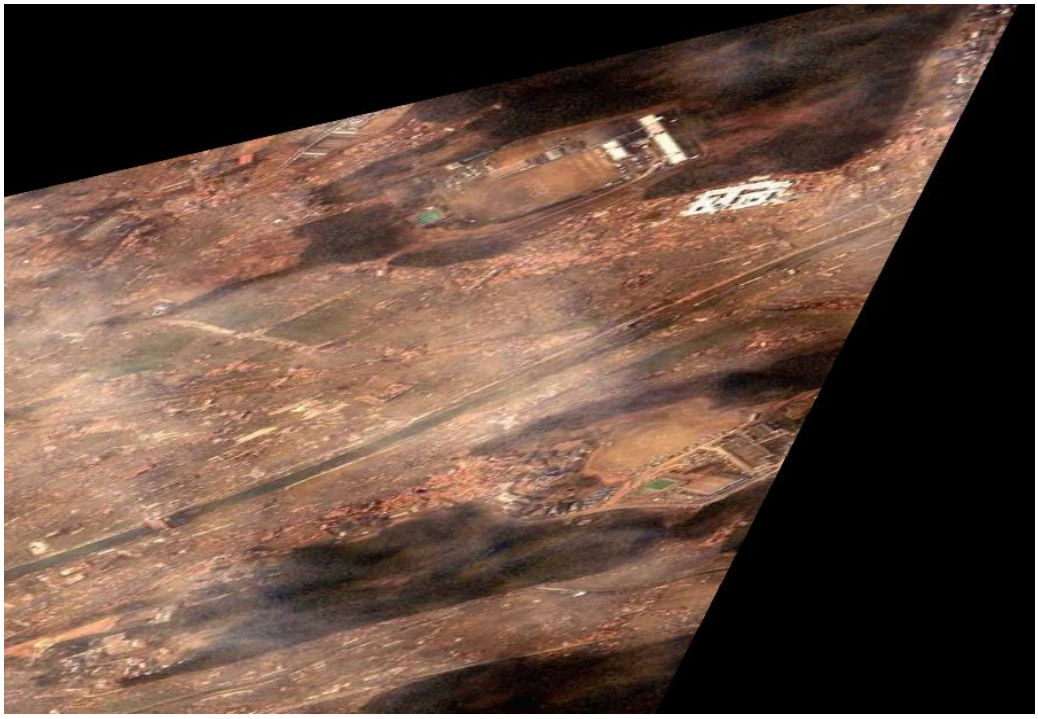
The subset of test images categorized as class 1 was registered using the limited window search approach with the proposed graph-based region descriptors. It is seen that this registration approach worked well in scenarios where content differences between the images are minimized. Such content differences could include object obstruction in an urban landscape or severe damage from a natural disaster. However, the limited window registration approach is shown to be robust in images with photometric variations such as differences in lighting conditions. Additionally, the class 1 image set provides evidence that the proposed registration technique is robust in scenarios with translation and rotational differences.

When compared to a common intensity-based method, the proposed method is shown to provide better registration where Fig. 6.16 illustrates one example. The average mutual information score reported in Table 6.1 also indicates the proposed limited window method is superior to the intensity-based method. Similarly, the registration results of the feature-based approach are shown to be inaccurate as compared to the proposed technique. For the example depicted in Fig. 6.16d, the set of matched features contained many outliers that adversely affected the parameter estimation, which can be attributed to the discriminative characteristics of the SURF descriptor in natural scenes.


a. Original reference image

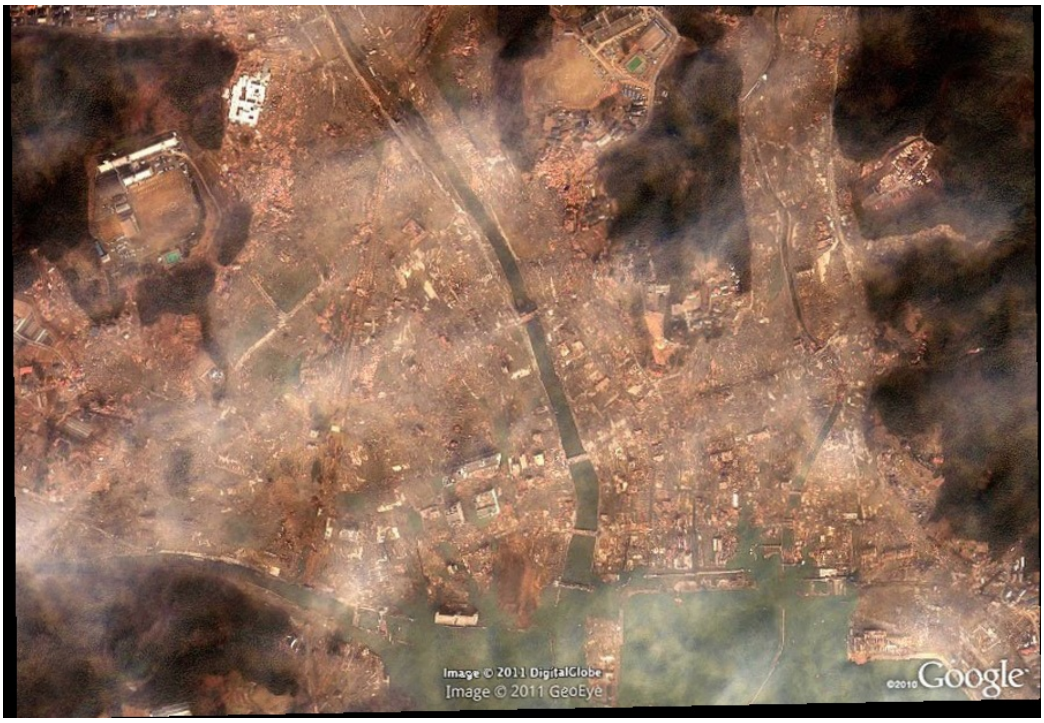



b. Original query image



c. Registered query image using an intensity-based approach

d. Registered query image using a feature-based approach




e. Registered query image using the limited window registration technique

Fig. 6.16 Example class 1 registration results comparing the proposed method to common intensity and feature-based methods

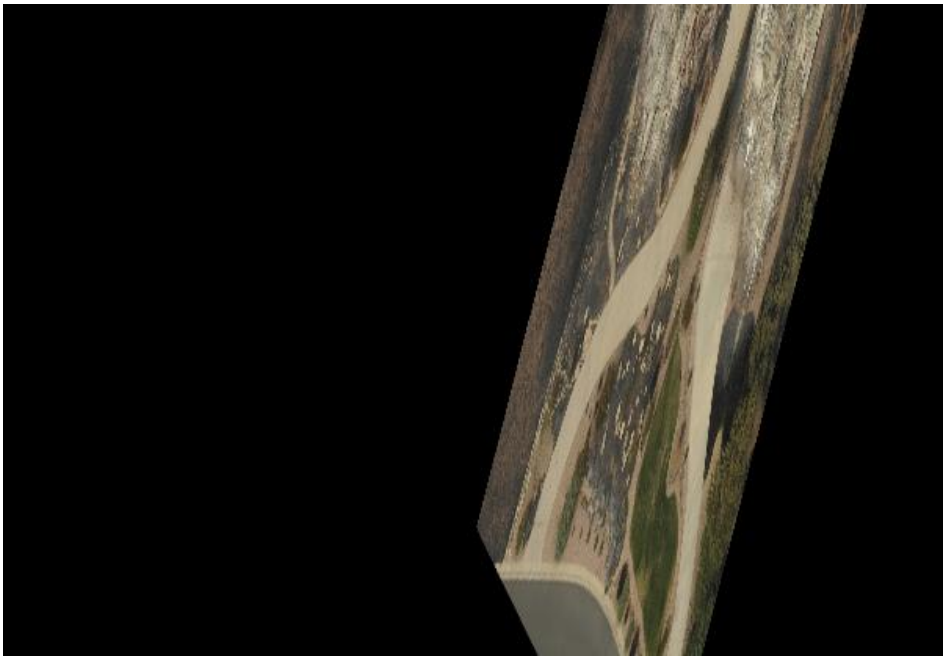
### 6.5.2. Class 2 Result Comparisons

Approximately 12.2% of the test set relied on the comprehensive registration technique. This class of images is observed to be robust in situations involving content differences and photometric variation; however, this proposed approach is adversely affected by large perspective and rotational differences. Because the proposed method first identifies the ten best similar image regions before symmetric feature matching, content differences between images are ignored. The ability for the method to disregard content differences increases the probability that the registration will be accurate. One example of the comprehensive approach outperforming typical intensity and feature-based approaches is given in Fig. 6.17 and Table 6.1.

a. Original reference image

b. Original query image

c. Registered query image using an intensity-based approach



d. Registered query image using a feature-based approach

e. Registered query image using the comprehensive registration technique

Fig. 6.17 Example class 2 registration results comparing the proposed method to common intensity and feature-based methods

### 6.5.3. Class 3 Result Comparisons

The set of class 3 images were found to rely on the exhaustive search registration technique with the proposed graph-based region descriptors. This method was employed if the coarse registration, limited window registration and comprehensive registration approaches were detected to be invalid through the proposed color histogram comparison method. This proposed registration method is observed to be robust in situations involving larger translation and rotational differences as well as photometric variations such as dissimilar illumination characteristics. The typical intensity and feature-based approaches exhibited lower mutual information scores for this class of images, as shown in Table 6.1, while Fig. 6.18 provides one example where the proposed technique outperforms the typical methods.

a. Original reference image

b. Original query image

c. Registered query image using an intensity-based approach

d. Registered query image using a feature-based approach

e. Registered query image using the exhaustive search registration with the proposed graph-based region descriptors

Fig. 6.18 Example class 3 registration results comparing the proposed method to common intensity and feature-based methods

### 6.6. GPU Implementation Considerations

The average execution times for the limited window region descriptor registration method, SURF registration with initial NCC coarse search, and unrestricted region descriptor registration approaches are 5.82s, 5.7s, and 10.6s, respectively. The K-means clustering and graph matching scheme incur the largest latencies in the sequential implementation. These tasks, along with feature point detection, Canny edge map creation, log-polar registration, and direct linear transform, can all benefit greatly from a parallel implementation on a GPU. From the results of recent literature in related applications [78-83], it can be deduced that the following potential speedups in Table 6.3 are feasible on a current NVidia GPU. The estimated log-polar registration speedup is

primarily determined by the reported speedup of the Canny edge detection on a GPU. Unfortunately the Canny edge detection speedup is not as efficient as the other tasks considered, however the average execution time for the sequential version of log polar registration is 82.3ms with a preprocessing step of down sampling.

Table 6.3 Estimated speedup factors on a GPU

| Task | Speedup |
|---|---|
| SURF | 46 |
| Clustering | 40 |
| Log Polar Registration | 3.43 |
| Graph Matching | 32 |
| Template Matching | 137.9 |

Using the estimated speedups, the inferred average execution times are calculated to be 172ms, 114ms, and 428ms, with overall speedups, 33.8, 50.9, and 24.77, respectively. A complete analysis of the proposed method's time complexity along with GPU time estimations can be found in Appendix A.

## 6.7. Summary

In this chapter, individual results are provided for the proposed coarse registration using the Canny edge maps, the limited window region descriptor registration approach, registration with SURF feature points utilizing normalized cross-correlation template matching, and unrestricted region descriptor matching registration. Overall results that use the mutual information metric show the viability of the proposed methods in situations with large photometric and geometric variations. Moreover, the proposed

method is shown to be effective in urban scenes with large obstructions. Lastly, the average execution time is discussed along with estimated speedups resulting in a GPU-based implementation.

# 7. CONCLUSIONS AND FUTURE WORK

## 7.1. Conclusions

In this thesis a novel multi-stage registration process is proposed which utilizes an effective graph-based region descriptor. The proposed approach attempts several methods for registration while evaluating each registration result between stages. This approach is shown to be a viable solution for registering images of scenes before and after a disaster. In such a scenario, the images to be registered may exhibit great variation in photometric and geometric characteristics.

The proposed method attempts an initial coarse registration, by estimating the translation and rotation parameters that relate two images, exploiting the shift properties of the image's Fourier transforms. The cross-correlation spectrum is calculated from the original images while rotation is estimated by first calculating the image's edge maps then converting the edge map images to the log-polar domain. The cross-correlation matrix derived from the log-polar edge maps is employed to estimate the rotation parameters. It is shown that the root mean square error (RMSE), calculated between the recovered rotation angle and expected angle, is lower when using the Canny edge maps and morphological operators for angles between $1^{\circ}$ and $180^{\circ}$. Furthermore, the RMSE is shown to be consistent when estimating the translation offsets, regardless of whether or not the edge maps are utilized. Although the utilization of edge maps improves the original phase correlation method, there still exists room to improve the registration. Experimental results show that an RMSE of $4^{\circ}$ is realized for angle differences of $180^{\circ}$. Similarly, the translational RMSE is shown to be approximately 30 pixels.

This work proposes an efficient measure for registration accuracy. The color histograms of the registered image are compared to the histograms of the original query image, where three comparisons are made, one for each of the color channels in the RGB color space. The hypothesis is that the color channel histograms are similar between the registered and original query images. This method is shown to be an effective determinant of successful registration and drives the proposed multi-stage registration process.

If the initial coarse registration step is determined to be successful, a novel proposed graph-based region descriptor is utilized within a limited search window to determine suitable control points for registration. The proposed region descriptor is extracted from the original reference image and the coarsely registered query image. These graphs are then matched across images, but within a specific boundary which is shown to greatly improve performance and accuracy.

The proposed region descriptor is structured from groups of invariant feature points. Two methods are proposed for combining the feature points. The first method considers the color characteristics as well as the spatial relationships between feature points where a two pass process is executed. The alternative method is a modified K-means clustering algorithm that is employed to collect the feature points based on their spatial relations. Due to the potential large variation in photometric properties of a scene or object from the effects of a natural disaster, reliance on color properties is not ideal; therefore the K-means approach is shown to be superior for the test cases. The additional computational complexity incurred by the K-means algorithm is justified through a higher rate of successful registrations.

Clusters of invariant feature points are used as the basis for creating the graph-based region descriptor, where the keypoints represent the nodes of the graph and the edges of the graph are determined using a shortest path algorithm. As a preprocessing step, feature points whose descriptors overlap are initially removed. Then the k-nearest

neighbor (k-NN) graph is structured as the shortest path from an initial node to all other nodes, where k = 1. The initial node is chosen as the node whose descriptor is most similar to the average descriptor value for all nodes in that particular cluster. Two graphs are compared using the proposed similarity metric that couples descriptor characteristics and spatial features. Graph structure is compared using the angles between consecutive edges and the angles between every other edge. These angles form the angle descriptor for the proposed region descriptor where comparisons are efficiently evaluated by converting the angles to the Binary Angle Measure (BAM) then encoding the BAM representation using the gray code. This form allows for a low latency assessment using the Hamilton distance.

Experimental results validate the matching ability of the proposed region descriptor. It is shown that the graph-based descriptor provided a higher matching rate than SIFT, SURF and BRISK for most scenarios in the test set. The proposed region descriptor is shown to be effective in scenarios that exhibit large variation in pixel intensities and structural differences. Moreover, the proposed descriptor is shown to identify image features within urban scenes and general objects.

When the coarse registration or limited window approach fails, the proposed method identifies similar image regions between the reference and query images. The normalized cross-correlation function acts as the basis for template matching where patches from the query image are compared to a sliding window in the reference image. The best matched image patches serve as the basis for invariant feature matching. The extracted feature points are used directly as control points for estimating the transformation parameters. For the given test set, 12.2% of the image pairs were registered using SURF feature points with the proposed initial coarse search using the normalized cross-correlation. The experimental results indicate the SURF-based approach produces accurate registrations when photometric variation is minimized.

The last stage of the proposed method attempts to identify registration control points from an unrestricted matching scheme utilizing the proposed graph-based region descriptor. An exhaustive search is performed to match image features from the region descriptor. It is shown that 22% of the test images failed through the SURF matching and therefore required the unrestricted approach. Of the scenarios that employ the unrestricted technique, 66% were registered accurately. The proposed approach is demonstrated to be effective in natural scenes, aerial images and urban scenarios.

The average latency of the proposed registration method is determined experimentally to be 5.82s for the limited window approach, 5.7s for the SURF and normalized cross-correlation technique, and 10.62s for the unrestricted region descriptor method. It has been shown that the main tasks of the proposed method can greatly benefit from a GPU implementation. With the aid of a CUDA-enabled GPU, it has been estimated that latencies are reduced to 172ms, 114ms, and 428ms, respectively.

Overall, the proposed registration technique is shown to improve registration accuracy when compared to traditional techniques in scenarios where large variations exist in pixel intensities, such as lightning conditions or damage from a natural disaster, as well as geometric differences, such as perspective variations. The mutual information metric was used to show experimentally that 92.68% of the query images were successfully registered.

## 7.2. Thesis Contributions

In this thesis, the following contributions are realized.

- Proposed an improved method for recovering registration parameters using the log-polar transform of edge map images in the frequency domain. We have shown that the Canny edge images coupled with morphological operators will improve the RMSE of the detected and expected angles.

- Proposed an effective and computationally simple method for evaluating the registration results after each stage. The color histograms of the query image before and after registration are shown to provide an indicator for registration accuracy.

- Introduced a novel graph-based region descriptor that is shown to be effective for identifying image features when geometric and photometric variations are present.

- Defined an effective similarity score for comparing the proposed region descriptors. The structural properties of the graphs are represented with a proposed angle descriptor, where the angles are denoted with bit strings that are encoded using the Gray Code convention.

- Outlined a multi-stage registration technique that is shown to be a viable registration solution for images depicting scenes before and after a disaster. The registration approach utilizes the graph-based region descriptor to identify control points, whereas SURF feature points are employed when the region descriptor registration is detected to produce sub-par results.

### 7.3. Future Work

In the proposed method, several areas could potentially benefit from alternative methods. First, the clustering algorithm can directly affect the overall results. In this work, two methods are explored; however many other techniques exist that may be effective for identifying groups of invariant feature points, such as DBSCAN and expectation-maximization (EM) with Gaussian mixture models. By using an alternative clustering approach, the invariant feature points may be grouped according to other properties which may improve the resulting graph structure. Furthermore, the proposed

method may benefit from researching other methods for graph creation. The proposed technique uses a simple nearest neighbor approach to create simple graphs, however there may exist additional graph structures which may be more discriminative.

The time analysis presented in Appendix A outlines the computational complexity of the proposed approach along with potential speedups realized from the utilization of a CUDA-enabled GPU. Future work should be focused on proving the estimated speedups on current GPU architectures. Additionally, a real-time implementation could be realized with further algorithm optimizations such as pipelining and data sharing on a parallel architecture.

### 7.4. Publications

- F. Bowen, J. Hu, and E. Y. Du, “A Multistage Approach for Image Registration,” IEEE Transactions on Cybernetics, Submitted 10/2013.

- F. Bowen, E. Y. Du, and J. Hu, “Comprehensive Feature and Texture Fusion-based Image Registration Approach,” Proceedings of SPIE, 2012.

- F. Bowen, E. Y. Du, and J. Hu, “A Novel Graph-based Invariant Region Descriptor for Image Matching,” IEEE International Conference on Electro/Information Technology (IEEE EIT), 2012.

- F. Bowen, E. Y. Du, J. Hu, “New Region Feature Descriptor-based Image Registration Method,” IEEE International Conference on Systems, Man, and Cybernetics, 2012.

# LIST OF REFERENCES

# LIST OF REFERENCES


[1] Chao Zhang, Chockalingam, P., Kumar, A., Burt, P., and Lakshmikumar, A., "Qualitative Assessment of Video Stabilization and Mosaicking Systems," *Applications of Computer Vision*, 2008. WACV 2008. IEEE Workshop on, vol., no., pp.1,6, 7-9 Jan. 2008.

[2] Fahmy, G., "Super-resolution construction of IRIS images from a visual low resolution face video," *Signal Processing and Its Applications*, 2007. ISSPA 2007. 9th International Symposium on, vol., no., pp.1,4, 12-15 Feb. 2007.

[3] Tsamoura, E. and Pitas, I., "Automatic Color Based Reassembly of Fragmented Images and Paintings," *Image Processing, IEEE Transactions* on, vol.19, no.3, pp.680,690, March 2010.

[4] Yimin Yang, Wenting Lu, Domack, J., Tao L, Shu-Ching Chen, Luis, S., and Navlakha, J.K., "MADIS: A Multimedia-Aided Disaster information Integration System for emergency management," *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, 2012 8th International Conference on , vol., no., pp.233,241, 14-17 Oct. 2012.

[5] Serpico, S.B., Dellepiane, S., Boni, G., Moser, G., Angiati, E., and Rudari, R., "Information Extraction From Remote Sensing Images for Flood Monitoring and Damage Evaluation," *Proceedings of the IEEE* , vol.100, no.10, pp.2946,2970, Oct. 2012.

[6] Hui Lin, Peijun Du, Weichang Zhao, Lianpeng Zhang, and Huasheng Sun, "Image registration based on corner detection and affine transformation," *Image and Signal Processing (CISP)*, 2010 3rd International Congress on , vol.5, no., pp.2184,2188, 16-18 Oct. 2010.

[7] Tzimiropoulos, G., Argyriou, V., Zafeiriou, S., and Stathaki, T., "Robust FFT-Based Scale-Invariant Image Registration with Image Gradients," *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, vol.32, no.10, pp.1899,1906, Oct. 2010.

[8] Ye, Peng, and Liu, Fang, "Combining local and non-local shape descriptor for multi-modal image registration," *Computer Science and Network Technology (ICCSNT)*, 2012 2nd International Conference on , vol., no., pp.2150,2153, 29-31 Dec. 2012.

[9] Pluim, J.P.W., Maintz, J.B.A., and Viergever, M.A., "Mutual-information-based registration of medical images: a survey," *Medical Imaging, IEEE Transactions on*, vol.22, no.8, pp.986,1004, Aug. 2003.

[10] Thomas, J., Kareem, A., and Bowyer, K., "Fast robust perspective transform estimation for automatic image registration in disaster response applications," *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, vol., no., pp.2190,2193, 22-27 July 2012.

[11] Bowen, F., Du, E.Y., and Jianghai Hu, "A novel graph-based invariant region descriptor for image matching," *Electro/Information Technology (EIT)*, 2012 IEEE International Conference on , vol., no., pp.1,6, 6-8 May 2012.

[12] Yuanhang Cheng, Dingyu Xue, Yuanhang Cheng, and Yanping Li, "A Fast Mosaic Approach for Remote Sensing Images," *Mechatronics and Automation*, 2007. ICMA 2007. International Conference on , vol., no., pp.2009,2013, 5-8 Aug. 2007.

[13] Yi Wang, Fevig, R., and Schultz, R.R., "Super-resolution mosaicking of UAV surveillance video," *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on* , vol., no., pp.345,348, 12-15 Oct. 2008.

[14] Rueckert, D., and Aljabar, P., "Nonrigid Registration of Medical Images: Theory, Methods, and Applications [Applications Corner]," *Signal Processing Magazine, IEEE* , vol.27, no.4, pp.113-119, July 2010.

[15] He, Y.P. and Gu, L.X., "Medical Image Registration Using Normal Vector and Intensity Value," *Intelligent Computation and Bio-Medical Instrumentation (ICBMI)*, 2011 International Conference on , vol., no., pp.46-49, 14-17 Dec. 2011.

[16] Zhi Li Song and Junping Zhang, "Remote Sensing Image Registration Based on Retrofitted SURF Algorithm and Trajectories Generated From Lissajous Figures," *Geoscience and Remote Sensing Letters, IEEE*, vol.7, no.3, pp.491-495, July 2010.

[17] Senarathne, C.N., Ransiri, S., Arangala, P., Balasooriya, A., and De Silva, C., "A faster image registration and stitching algorithm," *Industrial and Information Systems (ICIIS), 2011 6th IEEE International Conference on*, vol., no., pp.66-69, 16-19 Aug. 2011.

[18] Gang Hong and Yun Zhang, "Combination of feature-based and area-based image registration technique for high resolution remote sensing image," *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007*, IEEE International , vol., no., pp.377-380, 23-28 July 2007.

[19] Yijian Pei, Hao Wu, Jiang Yu, and Guanghui Cai, "Effective Image Registration based on Improved Harris Corner Detection," *Information Networking and Automation (ICINA), 2010 International Conference on*, vol.1, no., pp.V1-93-V1-96, 18-19 Oct. 2010.

[20] Orchard, J., "Globally Optimal Multimodal Rigid Registration: An Analytic Solution using Edge Information," *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol.1, no., pp.I-485-I-488, Sept. 16 2007-Oct. 19 2007.

[21] Xiaolong Dai and Khorram, S., "A feature-based image registration algorithm using improved chain-code representation combined with invariant moments," *Geoscience and Remote Sensing, IEEE Transactions on*, vol.37, no.5, pp.2351-2362, Sep 1999.

[22] Chen Xing and Peihua Qiu, "Intensity-Based Image Registration by Nonparametric Local Smoothing," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.33, no.10, pp.2081-2092, Oct. 2011.

[23] Omer, O.A. and Tanaka, T., "Robust image registration based on local standard deviation and image intensity," *Information, Communications & Signal Processing, 2007 6th International Conference on* , vol., no., pp.1-5, 10-13 Dec. 2007.

[24] Stamos, I. and Leordeanu, M., "Automated feature-based range registration of urban scenes of large scale," *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol.2, no., pp. II-555-Ii-561 vol.2, 18-20 June 2003.

[25] Bentoutou, Y., Taleb, N., Bounoua, A., Kpalma, K., and Ronsin, J., "Feature Based Registration of Satellite Images," *Digital Signal Processing, 2007 15th International Conference on*, vol., no., pp.419-422, 1-4 July 2007.

[26] Hui Guo, Chengqi Cheng, and Yubo Yang, "An automated registration of RS images based on SURF and piecewise linear transformation," *Environmental Science and Information Application Technology (ESIAT)*, 2010 International Conference on, vol.3, no., pp.133-136, 17-18 July 2010.

[27] Jana, R. and Ray, C., "Image Registration Using Object Shape's Chain Code," *Image and Signal Processing, 2009. CISP '09. 2nd International Congress on*, vol., no., pp.1-5, 17-19 Oct. 2009.

[28] Hui Lin, Peijun Du, Weichang Zhao, Lianpeng Zhang, and Huasheng Sun, "Image registration based on corner detection and affine transformation," *Image and Signal Processing (CISP), 2010 3rd International Congress on*, vol.5, no., pp.2184-2188, 16-18 Oct. 2010.

[29] Wang, A., Zhe Wang, Dan Lv, and Zhizhen Fang, "Research on a novel non-rigid registration for medical image based on SURF and APSO," *Image and Signal Processing (CISP), 2010 3rd International Congress on* , vol.6, no., pp.2628-2633, 16-18 Oct. 2010.

[30] Khan, M.K. and Nyström, I., "A Modified Particle Swarm Optimization Applied in Image Registration," *Pattern Recognition (ICPR), 2010 20th International Conference on*, vol., no., pp.2302-2305, 23-26 Aug. 2010.

[31] Xinting Gao, Sattar, F., and Venkateswarlu, R., "Multiscale Corner Detection of Gray Level Images Based on Log-Gabor Wavelet Transform," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol.17, no.7, pp.868,875, July 2007.

[32] Luo Juan and Oubong Gwun, "SURF applied in panorama image stitching," *Image Processing Theory Tools and Applications (IPTA), 2010 2nd International Conference on*, vol., no., pp.495,499, 7-10 July 2010.

[33] Kennedy, J. and Eberhart, R., "Particle swarm optimization," *Neural Networks, 1995. Proceedings., IEEE International Conference on*, vol.4, no., pp.1942-1948 vol.4, Nov/Dec 1995.

[34] Myung-Eun Lee, Soo-Hyung Kim, and In-Hye Seo , "Intensity-based registration of medical images," *Test and Measurement, 2009. ICTM '09. International Conference on*, vol.1, no., pp.239-242, 5-6 Dec. 2009.

[35] Klein, S., Staring, M., Murphy, K., Viergever, M.A., and Pluim, J.P.W., "elastix: A Toolbox for Intensity-Based Medical Image Registration," *Medical Imaging, IEEE Transactions on*, vol.29, no.1, pp.196,205, Jan. 2010.

[36] Shams, R., Kennedy, R.A., Sadeghi, P., and Hartley, R., "Gradient Intensity-Based Registration of Multi-Modal Images of the Brain," *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, vol., no., pp.1-8, 14-21 Oct. 2007.

[37] Xiuqiong Zhang, "A Novel Quality Metric for Image Fusion Based on Color and Structural Similarity," *2009 International Conference on Signal Processing Systems*, vol., no., pp.258-262, 15-17 May 2009.

[38] Liang Lei, Tongqing Wang, Jun Peng, and Bo Yang, "Image retrieval based on intrinsic dimension and Shannon entropy," *Cognitive Informatics & Cognitive Computing (ICCI*CC ), 2011 10th IEEE International Conference on*, vol., no., pp.216-222, 18-20 Aug. 2011.

[39] Xiaoxin Guo, Zhiwen Xu, Yinan Lu, and Yunjie Pang, "An Application of Fourier-Mellin Transform in Image Registration," *Computer and Information Technology, 2005. CIT 2005. The Fifth International Conference on*, vol., no., pp.619,623, 21-23 Sept. 2005.

[40] Hanzhou Liu, Baolong Guo, and Zongzhe Feng, "Pseudo-log-polar Fourier transform for image registration," *Signal Processing Letters, IEEE*, vol.13, no.1, pp.17,20, Jan. 2006.

[41] Wolberg, G. and Zokai, S., "Robust image registration using log-polar transform," *Image Processing, 2000. Proceedings. 2000 International Conference on*, vol.1, no., pp.493,496 vol.1, 2000.

[42] Makwana, Y.N. and Somkuwar, A.K., "A Novel Technique for Robust Image Registration Using Log Polar Transform," *Communication Systems and Network Technologies (CSNT), 2012 International Conference on*, vol., no., pp.157,160, 11-13 May 2012.

[43] Shah, C.A., Yongwei Sheng, and Smith, L.C., "Automated Image Registration Based on Pseudoinvariant Metrics of Dynamic Land-Surface Features," *Geoscience and Remote Sensing, IEEE Transactions on*, vol.46, no.11, pp.3908,3916, Nov. 2008.

[44] D. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, *Int’l J. Computer Vision*, vol. 2, no.60, 2004, p91-110.

[45] Yan Ke and Sukthankar, R., "PCA-SIFT: a more distinctive representation for local image descriptors," *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol.2, no., pp. II-506- II-513 Vol.2, 27 June-2 July 2004.

[46] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool, “Speeded-Up Robust Features (SURF),” *International Journal of Computer Vision and Image Understanding (CVIU)*, Vol. 110, No. 3, pp. 346-359, 2008.

[47] Zen Chen and Shu-Kuo Sun, "A Zernike Moment Phase-Based Descriptor for Local Image Representation and Matching," *Image Processing, IEEE Transactions on*, vol.19, no.1, pp.205-219, Jan. 2010.

[48] Doretto, G. and Yi Yao, "Region moments: Fast invariant descriptors for detecting small image structures," *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, vol., no., pp.3019-3026, 13-18 June 2010.

[49] Dinesh Kumar, V.P. and Thomas, T., "Clustering of Invariance Improved Legendre Moment Descriptor for Content Based Image Retrieval," *Signal Processing, Communications and Networking, 2008. ICSCN '08. International Conference on*, vol., no., pp.323-327, 4-6 Jan. 2008.

[50] Leutenegger, S., Chli, M., and Siegwart, R.Y., "BRISK: Binary Robust invariant scalable keypoints," *Computer Vision (ICCV), 2011 IEEE International Conference on*, vol., no., pp.2548,2555, 6-13 Nov. 2011.

[51] Dawei Wang, Wei Ge, and Yanjie Wang, "Using bidirectional Binary Particle Swarm Optimization for feature selection in feature-level fusion recognition system," *Industrial Electronics and Applications, 2009. ICIEA 2009. 4th IEEE Conference on*, vol., no., pp.3810-3814, 25-27 May 2009.

[52] Awais, M. and Mikolajczyk, K., "Feature Pairs Connected by Lines for Object Recognition," *Pattern Recognition (ICPR), 2010 20th International Conference on*, vol., no., pp.3093-3096, 23-26 Aug. 2010.

[53] Hong-Min Liu, Zhi-Heng Wang, and Chao Deng, "Extend point descriptors for line, curve and region matching," *Machine Learning and Cybernetics (ICMLC), 2010 International Conference on*, vol.1, no., pp.214-219, 11-14 July 2010.

[54] Harris C. and Stephens M.A., “Combined Corner and Edge Detector,” *Proceedings of the Fourth Alvey Vision Conference*, Manchester, UK, 1988:147-151.

[55] Azad, P., Asfour, T., and Dillmann, R., "Combining Harris interest points and the SIFT descriptor for fast scale-invariant object recognition," *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, vol., no., pp.4275,4280, 10-15 Oct. 2009.

[56] Jieyu Zhang, Qiang Chen, Xiaojing Bai, Quansen Sun, Huaijiang Sun, and Deshen Xia, "An Advanced Harris-Laplace Feature Detector with High Repeatability," *Image and Signal Processing, 2009. CISP '09. 2nd International Congress on*, vol., no., pp.1,5, 17-19 Oct. 2009.

[57] Guest, R. and Miguel-Hurtado, O., "Enhancing static biometric signature verification using Speeded-Up Robust Features," *Security Technology (ICCST), 2012 IEEE International Carnahan Conference on*, vol., no., pp.213,217, 15-18 Oct. 2012.

[58] Pinto, B. and Anurenjan, P. R., "Video stabilization using Speeded Up Robust Features," *Communications and Signal Processing (ICCSP), 2011 International Conference on*, vol., no., pp.527,531, 10-12 Feb. 2011.

[59] Bin Li, Maohua Wang, and Li Li, "A Real-Time Pineapple Matching System Based on Speeded-Up Robust Features," *Computational Intelligence and Security (CIS), 2010 International Conference on*, vol., no., pp.243,247, 11-14 Dec. 2010.

[60] Zhijie Zhang, Chongxiao Cao, Ruijie Zhang, and Jianhua Zou, "Video copy detection based on Speeded Up Robust Features and Locality Sensitive Hashing," *Automation and Logistics (ICAL), 2010 IEEE International Conference on*, vol., no., pp.13,18, 16-20 Aug. 2010.

[61] Sirmacek, B. and Unsalan, C., "Urban-Area and Building Detection Using SIFT Keypoints and Graph Theory," *Geoscience and Remote Sensing, IEEE Transactions on*, vol.47, no.4, pp.1156-1167, April 2009.

[62] Kisku, D.R., Rattani, A., Grosso, E., and Tistarelli, M., "Face Identification by SIFT-based Complete Graph Topology," *Automatic Identification Advanced Technologies, 2007 IEEE Workshop on*, vol., no., pp.63-68, 7-8 June 2007.

[63] Charalampidis, D., "A modified k-means algorithm for circular invariant clustering," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.27, no.12, pp.1856,1865, Dec. 2005.

[64] Jie Zhao, Hui-Juan Zhou, and Guo-Zun Men, "A method of sift feature points matching for image mosaic," *Machine Learning and Cybernetics, 2009 International Conference on*, vol.4, no., pp.2353,2357, 12-15 July 2009.

[65] MacQueen, J. B., “Some Methods for Classification and Analysis of MultiVariate Observations,” *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp.281-297, University of California Press, Year 1967.

[66] Bowen, F., Du, E., Jianghai Hu, "New region feature descriptor-based image registration method," *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*, vol., no., pp.2489,2494, 14-17 Oct. 2012.

[67] Foss, F.A., "The use of a reflected code in digital control systems," *Electronic Computers, Transactions of the I.R.E. Professional Group on*, vol.EC-3, no.4, pp.1,6, Dec. 1954

[68] Gray, F. "Pulse Code Communication." *United States Patent Number 2632058*, March 17, 1953.

[69] Hamming, R.W., “Error Detecting and Error Correcting Codes,” *Bell System Technical Journal*, (29) 2: 147--160, Amer Telephone Telegraph CO, NEW YORK, 1950.

[70] Hartley, R. and Zisserman, A., “Multiple View Geometry in Computer Vision,” Cambridge University Press, 2000.

[71] Teke, M. and Temizel, A., "Multi-spectral Satellite Image Registration Using Scale-Restricted SURF," *Pattern Recognition (ICPR), 2010 20th International Conference on*, vol., no., pp.2310,2313, 23-26 Aug. 2010.

[72] Amsterdam Library of Object Images, *http://staff.science.uva.nl/~aloi/*.

[73] Canny, John, "A Computational Approach to Edge Detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.PAMI-8, no.6, pp.679-698, Nov. 1986.

[74] Kanopoulos, N., Vasanthavada, N., and Baker, R.L., "Design of an image edge detection filter using the Sobel operator," *Solid-State Circuits, IEEE Journal of*, vol.23, no.2, pp.358,367, Apr 1988.

[75] Hsu-Yung Cheng, Chih-Chia Weng, and Yi-Ying Chen, "Vehicle Detection in Aerial Surveillance Using Dynamic Bayesian Networks," *Image Processing, IEEE Transactions on*, vol.21, no.4, pp.2152,2159, April 2012.

[76] J. Serra, “Image Analysis and Mathematical Morphology,” London, U.K.: Academic, 1982.

[77] L. Najman and H. Talbot, “Mathematical Morphology: From Theory to Applications,” New York: ISTE-Wiley, 2010.

[78] Yuancheng Luo and Duraiswami, R., "Canny edge detection on NVIDIA CUDA," *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*, vol., no., pp.1,8, 23-28 June 2008.

[79] Zechner, M. and Granitzer, M., "Accelerating K-Means on the Graphics Processor via CUDA," *Intensive Applications and Services, 2009. INTENSIVE '09. First International Conference on*, vol., no., pp.7,15, 20-25 April 2009.

[80] Bai Hong-tao, He Li-li, Ouyang Dan-tong, Li Zhan-shan, and Li He, "K-Means on Commodity GPUs with CUDA," *Computer Science and Information Engineering, 2009 WRI World Congress on*, vol.3, no., pp.651,655, March 31 2009-April 2 2009.

[81] Junchul Kim, Eunsoo Park, Xuenan Cui, Hakil Kim, and Gruver, W.A., "A fast feature extraction in object recognition using parallel processing on CPU and GPU," *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, vol., no., pp.3842,3847, 11-14 Oct. 2009.

[82] Kauker, D., Sanftmann, H., Frey, S., and Ertl, T., "Memory Saving Discrete Fourier Transform on GPUs," *Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on*, vol., no., pp.1152,1157, June 29 2010-July 1 2010.

[83] Zhenman Fang, Donglei Yang, Weihua Zhang, Haibo Chen, and Binyu Zang, "A comprehensive analysis and parallelization of an image retrieval algorithm," *Performance Analysis of Systems and Software (ISPASS), 2011 IEEE International Symposium on*, vol., no., pp.154,164, 10-12 April 2011.

[84] Kai Briechle and Uwe D. Hanebeck, "Template matching using fast normalized cross correlation," *Proc. SPIE 4387, Optical Pattern Recognition XII*, 95 (March 20, 2001).

[85] De Beer, R., Van Ormondt, D., Di Cesare, F., Graveron-Demilly, D., Karras, D.A., and Starcuk, Z., "Accelerating batched 1D-FFT with a CUDA-capable computer," *Imaging Systems and Techniques (IST), 2010 IEEE International Conference on*, vol., no., pp.446,451, 1-2 July 2010.

[86] Nishida, K., Ito, Y., and Nakano, K., "Accelerating the Dynamic Programming for the Matrix Chain Product on the GPU," *Networking and Computing (ICNC), 2011 Second International Conference on*, vol., no., pp.320,326, Nov. 30 2011-Dec. 2 2011.

[87] Nan Zhang, "Computing Optimised Parallel Speeded-Up Robust Features (P-SURF) on Multi-Core Processors," *International Journal of Parallel Programming*, volume 38, number 2, pages 138-158, 2010.

[88] Inaba, M., Katoh, N., and Imai, H., "Applications of weighted Voronoi diagrams and randomization to variance-based k-clustering". *Proceedings of 10th ACM Symposium on Computational Geometry*, pp. 332–339, 1994.

[89] Pawan Harish and P. J. Narayanan, “Accelerating large graph algorithms on the GPU using CUDA,” *In Proceedings of the 14th international conference on High performance computing (HiPC'07)*, Springer-Verlag, Berlin, Heidelberg, 197-208, 2007.

[90] Lahabar, S. and Narayanan, P. J., "Singular value decomposition on GPU using CUDA," *Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on*, vol., no., pp.1,10, 23-29 May 2009.

# APPENDIX

# A. TIME ANALYSIS AND PROPOSED GPU IMPLEMENTATION

The following chapter provides a thorough analysis of the proposed algorithm's complexity and provides an in-depth discussion about GPU implementations of the tasks that are utilized in the proposed method. In subsequent sections, each major task of the proposed approach is examined to provide a proposed GPU implementation where previous literature is used to support the proposal. GPU performance will vary based on architecture and the algorithm itself; however the proposed registration technique contains many areas where significant speedup can be realized with a parallel or distributed implementation. Prior to the complexity analysis, a brief introduction on CUDA is presented.

## A.1 CUDA

The Compute Unified Device Architecture (CUDA) is a hardware platform and tightly couple software API that allows an application to take advantage an Nvidia GPU's set of single-instruction/multiple-data (SIMD) processors. Current GPUs contain up to 32 SIMD processors where each processor is coupled with a set of stream processors. For a given clock cycle, all grouped stream processors execute the same instruction, but with different data. This architecture is illustrated in Fig. A.1.

The CUDA platform offers several levels of parallelism. True thread level parallelism can be achieved by distributing the kernels to different SIMD processors. Multiple threads can be assigned to a single SIMD, however those threads are time-sliced and share the SIMD resources. Blocks of these time-sliced threads on a SIMD are called

warps. Nvidia GPUs also provide instruction level parallelism through the concurrent instruction execution on the stream processors.

Applications interface to CUDA-capable GPUs through the CPU's main memory and the GPU's shared memory. Although there is ample opportunity to exploit parallel execution, there exists a nontrivial latency when transferring data to and from the GPU for execution. Because of this fact, CUDA implementations exhibit the largest speedups in performance when large datasets are used that require many repetitions of the same set of instructions. Moreover, the amount of parallelism relies upon the algorithm and its associated data dependencies. In the following sections, the computational complexity of the major tasks of the proposed registration algorithm is analyzed to determine the effectiveness of GPU execution.

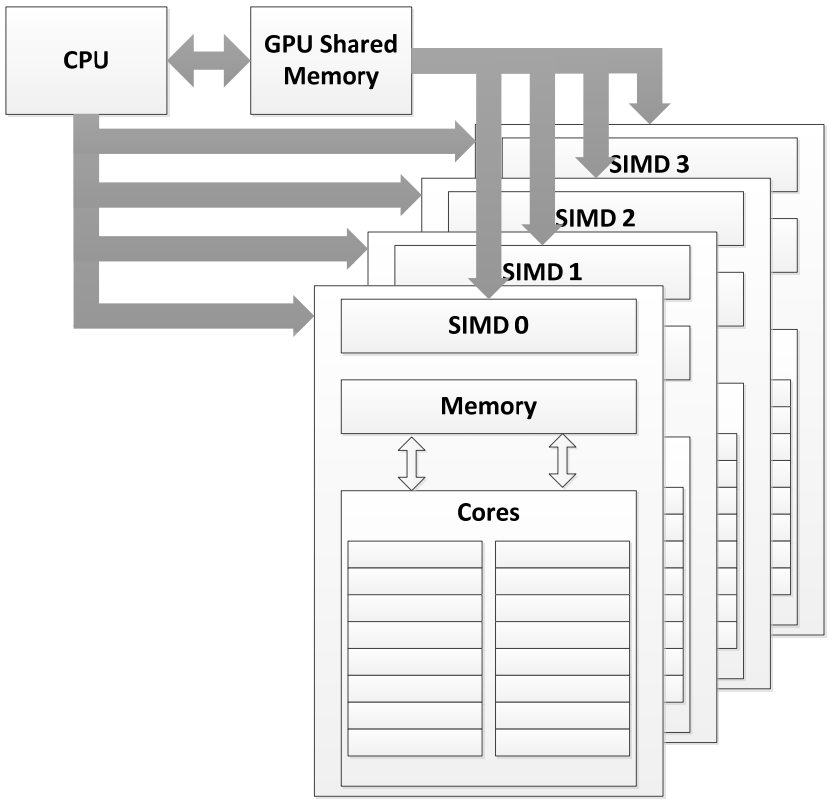



Fig. A.1 Nvidia CUDA hardware architecture

## A.2 Coarse Registration

### A.2.1 Time Complexity

The proposed coarse registration technique involves converting an image to log polar coordinates, transferring the images to the frequency domain, performing matrix multiplication, transferring the result to the time domain, and exhaustively searching for a peak within the cross-correlation matrix. Given an image, $I(x,y) \in \mathbb{R}^{M\ x\ N}$, the complexity of the log polar transform is $O(MN)$. The forward discrete Fourier transform (DFT) is defined in (A.1) whereas the inverse discrete Fourier transform (IDFT) is given in (A.2).

$$I(u,v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} I(x,y)e^{-j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)} \tag{A.1}$$

$$I(x,y) = \sum_{u=0}^{M-1}\sum_{v=0}^{N-1} I(u,v)e^{j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)} \tag{A.2}$$

For both cases, DFT and IDFT, the computational complexity of a naïve implementation is $O((MN)^2)$, while the matrix multiplication used for computing the power spectrum cross-correlation is $O(M^2N)$.

Canny edge detection can be viewed as a three phase method. First, the initial Gaussian smoothing consists of $25MN$ computations for a $5\ x\ 5$ window size and an $M\ x\ N$ image. The gradient detection using $3\ x\ 3$ Sobel filters requires $50MN$ operations while the calculation of gradient magnitude and orientation require $2MN$ operations. Lastly, non-maximal suppression and hysteresis require $3MN$ operations; therefore the overall complexity of the sequential canny edge detection is shown to be,

$$O(25MN) + O(50MN) + O(2MN) + O(3MN) \leq O(MN) \tag{A.3}$$

The complexity for the entire coarse registration process is derived as follows,

$$\begin{aligned} &O(MN) + O(2 * (MN)^2) + O(M^2N) + O(MN) \\ &\leq O(M^2N^2) + O\big(N(M^2 + M)\big) \\ &\leq O(M^2N^2) + O(M^2N^2) \\ &\leq O(M^2N^2) \end{aligned} \tag{A.4}$$

### A.2.2 GPU Time Complexity

GPU implementations for the Fast Fourier Transform (FFT), inverse Fast Fourier Transform (IFFT), and matrix multiplication have been research extensively. The Fast Fourier Transform (FFT) exploits the transform's ability to separate the transform into a series of 1-D calculations. In such a scenario, the computational complexity is shown to be $O(MN \log MN)$. In [82], Kauker *et al*, propose a memory efficient DFT implementation for GPUs which reduces the complexity to $O\left(M(\frac{N}{2}+1)\log M(\frac{N}{2}+1)\right)$, while the authors of [85] report a speedup of 20 times faster than sequential FFT implementations. Matrix multiplication is shown to be improved by a factor of 40 when utilizing linear programming methodologies on a GPU [86]. Luo and Duraiswami propose a CUDA implementation of a Canny edge detector that achieves a 3.403 speedup over an optimized sequential OpenCV implementation [78].

## A.3 SURF

### A.3.1 SURF Complexity

Speeded up Robust Features is an ubiquitous feature point detection and description scheme that aimed to provide the same level of accuracy and robustness over SIFT while greatly reducing the computational complexity through the use of box filters and integral

images. The following complexity derivation assumes an image is defined as, $I(x, y) \in \mathbb{R}^{M\,x\,N}$, and the number of detected feature points is represented by $k$. Table A.1 summarizes the primary tasks of SURF along with the number of operations per task. The number of operations is listed assuming the use of 4 octaves with 4 scales which are the preferred parameters by Bay *et al* [46].

Table A.1 Number of computations required for SURF feature point detection and descriptor extraction

| SURF Task | Number of Operations | Notes |
|---|---|---|
| Create integral image | MN | |
| Computing Hessian matrix for all octaves using box filters and the integral image | 512MN | 4 scales at 4 octaves with 26 additions and 6 multiplications |
| Non-maximal suppression for feature selection | 16MN | 4 scales at 4 octaves |
| Orientation assignment | $36s^2$k | *s* denotes the detected scale for a particular feature point. Orientation assignment occurs within a radius 6s. |
| Descriptor extraction | 208k | 4x4 grid about the feature point with 9 multiplications and 4 additions |

$$O(MN) + O(512MN) + O(16MN) + O(36s^2k) + O(208k) \leq O(MN + s^2k) \quad \text{(A.5)}$$

Traditional approaches to SURF feature point matching involve an exhaustive search for the best matched pairs while instituting a nearest-neighbor ratio rule. The NN ratio compares the best matched feature point and the second best matched feature point. Smaller ratios indicate a higher confidence in the match. In the proposed comprehensive registration approach, invariant feature points are matched symmetrically also coupled with the NN ratio. Let $k_r$ and $k_q$ represent the number of detected feature points in the reference and query images, respectively, and define the number of feature points as $k = min(k_r, k_q)$. The complexity of the symmetric feature point matching is then given as,

$$O(2k^2) \leq O(k^2) \tag{A.6}$$

### A.3.2 SURF GPU

Although the original SURF was able to achieve massive performance gains over SIFT, there was still a need to further decrease the algorithm's latency. There have been several versions of SURF proposed that improve execution time, including P-SURF [87], and CUDA SURF [81]. Fang *et al* [83] proposed a SURF implementation for GPUs that utilizes block parallelism as well as pipelining, which is shown to offer a speedup factor of 46. This architecture is shown to outperform CUDA SURF by a factor of 1.5.

## A.4 Template Matching

In the proposed registration method, a comprehensive registration method is attempted if the initial limited window graph-based region descriptor approach fails. The first step is to identify similar image patches between the reference and query images. For the proposed approach, normalized cross-correlation template matching is exploited. Given a template from the query image, $t \in \mathbb{R}^{m\,x\,n}$, and the reference image, $I_R \in \mathbb{R}^{M\,x\,N}$, the normalized cross-correlation score at location $(u, v)$ is defined as,

$$NCC = \frac{\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(I_R(x,y) - \bar{I_R})(t(x-u,y-v) - \bar{t})}{\sqrt{\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(I_R(x,y) - \bar{I_R})^2 \sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(t(x-u,y-v) - \bar{t})^2}}. \quad \text{(A.7)}$$

$\bar{I_R}$ and $\bar{t}$ denote the average pixel values of the reference image within the current search window and average pixel value within the given template, respectively. The complexity of matching a single template is defined as,

$$O\big(mn(M-m+1)(N-n+1)\big). \quad \text{(A.8)}$$

In the proposed approach, the query image is segmented into $50\,x\,50$ templates. Each template is compared to the reference image while the best 10 matched templates are used in subsequent steps, therefore the complexity of the proposed template matching scheme is,

$$\begin{aligned} &O\left(mn(M-m+1)(N-n+1)\left(\frac{M}{50}\right)\left(\frac{N}{50}\right)\right) \\ &\leq O\big(mnMN(M-m+1)(N-n+1)\big) \\ &\leq O((mnMN)^2) \end{aligned} \quad \text{(A.9)}$$

In the method proposed by Gupta *et al*, a CUDA implementation of the fast normalized cross correlation template matching technique is shown to achieve a total speedup of 137.9 over a single threaded implementation.

### A.5 Graph Creation

K-means clustering is an effective technique that is exploited when creating the proposed graph-based region descriptor. Given the number of feature points, $n$, the number of clusters, $k$, and the dimension of each feature point, $d$, the complexity is derived to be [88],

$$O(n^{dk+1}\log n). \tag{A.10}$$

Due to the popularity of the algorithm and the opportunity for parallelization, several CUDA implementations for K-means clustering have been proposed [79,80]. Bai *et al* [80] proposed an implementation that achieves a speedup factor of 40 when the number of clusters is approximately 200 which is a comparable quantity to the number of graphs produced in the proposed method. The proposed use of k-means on a GPU is further justified when the detected feature points are previously stored in the GPU's memory.

Dijkstra's shortest path algorithm is the basis for constructing the proposed region descriptor. Let $m$ represent the number of nodes in a cluster after feature point pruning, then the complexity of Dijkstra's algorithm applied to the cluster of invariant feature points is defined as,

$$\begin{aligned} &O(8(m-1+m\log m)) \\ &\leq O(m+m\log m), \end{aligned} \tag{A.11}$$

where 8 operations are required to compare two nodes. For an image with $k$ clusters, the overall complexity for creating the graphs for an entire image is given as,

$$O(km+km\log m). \tag{A.12}$$

Harish and Narayanan [89] propose a CUDA implementation of Dijkstra's algorithm that achieves a speedup factor of 70, however their proposed method is tailored for large datasets. In the proposed method, there exists a large quantity of small datasets.

Graph creation is a highly parallelizable process that can benefit greatly from a CUDA implementation. Assuming the results from the k-means clustering is stored in the GPU's shared memory, a speedup of 16 is conceivable for an architecture with 32

SIMD processors. The overhead of transferring data to and from the host CPU can be neglected if the processed data originates from a previous set of GPU threads.

### A.6 Graph Matching

In both the limited neighborhood and unrestricted region descriptor registration methods, control points are identified as the centroids of two matched graphs. The proposed graph matching scheme is summarized in Section 3.3.2.4 where the following similarity score is used to quantify the overall correspondence of both structure and texture characteristics.

$$S = \frac{\sum_{|V|} S_{FP}}{|V|} + \frac{\sum_{k=1}^{|d_{\theta 1}^{\alpha}|} \left[ \frac{d_H\left(d_{\theta 1}^{\alpha}(k), d_{\theta 1}^{\rho}(k)\right)}{2^n} \right]}{\left|d_{\theta 1}^{\alpha}\right|} + \frac{\sum_{k=1}^{|d_{\theta 2}^{\alpha}|} \left[ \frac{d_H\left(d_{\theta 2}^{\alpha}(k), d_{\theta 2}^{\rho}(k)\right)}{2^n} \right]}{\left|d_{\theta 2}^{\alpha}\right|} \quad \text{(A.13)}$$

The complexity of matching $G_1$ and $G_2$ is determined as follows. Let $m = |G_1|$ and $n = |G_2|$. Then the total number of sub-graph comparisons is given by $m - n + 1$, where $m \geq n$. For each sub-graph comparison, there are $64n + 8(n-1) + 8(n-2) \leq O(n)$ operations; therefore the computational complexity of comparing two graphs is,

$$\begin{aligned} &O\big(n * (m - n + 1)\big) \\ &\leq O(nm) - O(n^2) + O(1) \\ &\leq O(nm) - O(n^2) \\ &\leq O(nm) \end{aligned} \quad \text{(A.14)}$$

For a given set of images with $M$ graphs in the one image and $N$ graphs in another, the computational complexity is given by,

$$O(MNnm). \tag{A.15}$$

Similar to graph creation, the latency of graph matching can be greatly reduced while exploiting the parallelism of a GPU. If the comparison to two graphs is completed by a single GPU SIMD processor, the node comparisons and angle comparisons can be completed in parallel on the core processors. As an example, if a similarity is calculated for two graphs, each with 5 nodes, then the similarity score will require 12 operations (5 node comparisons, 4 angle comparisons, and 3 secondary angle comparisons), which can easily be accomplished utilizing the parallel core processors. If SURF feature points are used, each node comparison entails 128 additions and 64 multiplications for computing the Euclidean distance of SURF descriptors. While executed in parallel on the stream processors, the Euclidean distance can be completed in constant time, $O(1)$. A conservative speedup factor of 32 is reasonable because the complexity when exploiting the GPU core processors is then

$$O(MN). \tag{A.16}$$

### A.7 Direct Linear Transform

The direct linear transform is utilized to estimate the transformation parameters that relate two images. The basis of this method relies on the singular value decomposition (SVD) of a set of linear equations that are constructed from a set of control points. The computational complexity of SVD for a $m\ x\ n$ matrix is derived by Golub and Van Loan in [90], where they state the complexity is,

$$O(m^2 + n^3). \tag{A.17}$$

Although DLT has a large computational complexity, the speedup advantages of a GPU implementation can only be realized with large datasets. The process of transferring the data and instructions to and from the GPU, as well as the computation

itself, incurs latencies which may be larger than completing the computation on the CPU. Such a scenario is reported by Lahabar and Narayanan where their CUDA implementation performed worse than a single threaded CPU for matrix sizes smaller than $750\ x\ 750$ [90].

### A.7 Overall Time Complexity of the Proposed Multistage Registration Method

The computational complexity of the proposed registration approach is provided in Table A.2 for *M x N* images with *m* detected graphs in one image and *n* detected graphs in another.

Table A.2 Complexity summary of the multistage registration approach

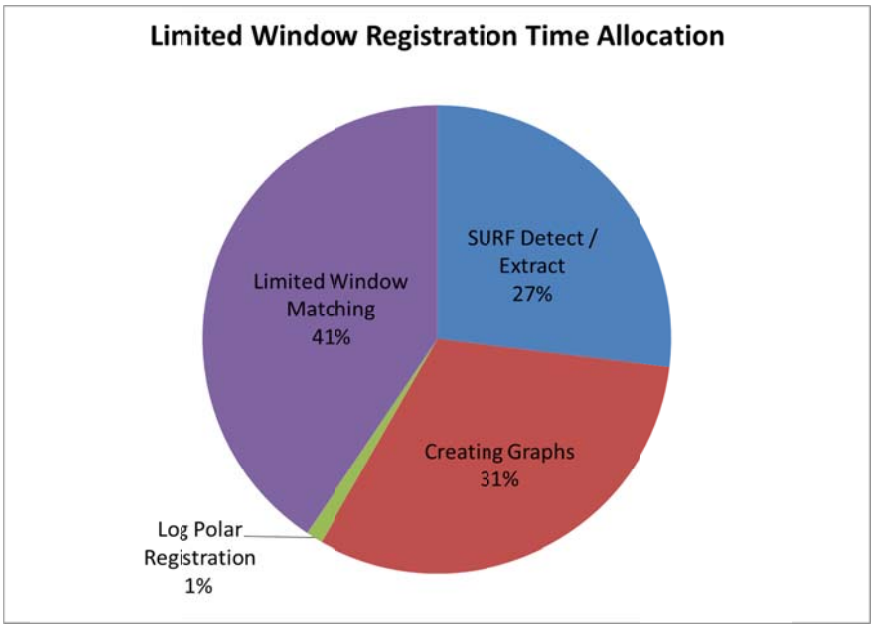
| Coarse Registration | SURF Detection | Graph Creation | Template Matching | Graph Matching | Direct Linear Transform |
|---|---|---|---|---|---|
| $M^2N^2$ | $MN + s^2k$ | $km + km\log m$ | $(mnMN)^2$ | $MNnm$ | $m^2 + n^3$ |

### A.8 Time Analysis and Theoretical CUDA Execution Time

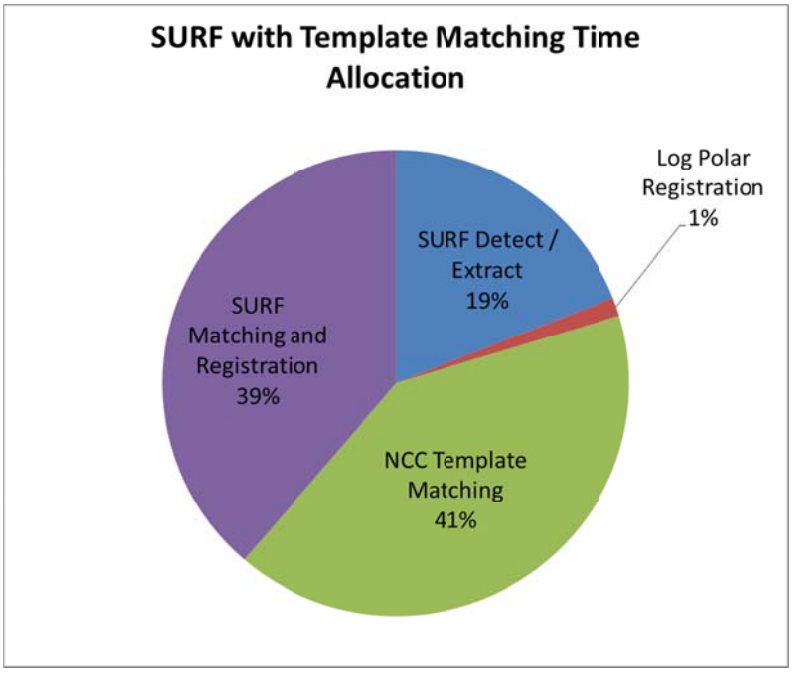
For the database of test images, the overall time allocation for each phase of the proposed method was determined. Each phase of each stage was timed to determine the overall time spent. Fig. A.2 provides illustrations summarizing the time allocations where it is evident that the most allocation occurs during the initial coarse registration. Table A.3 summarizes the average execution times for each phase of each stage. The stages of the proposed method, limited window region descriptor registration, SURF with coarse template matching, and unrestricted region descriptor registration, exhibit average execution times of 5.82s, 5.7s, and 10.6s, respectively.

Table A.3 Average execution times for each phase of each stage

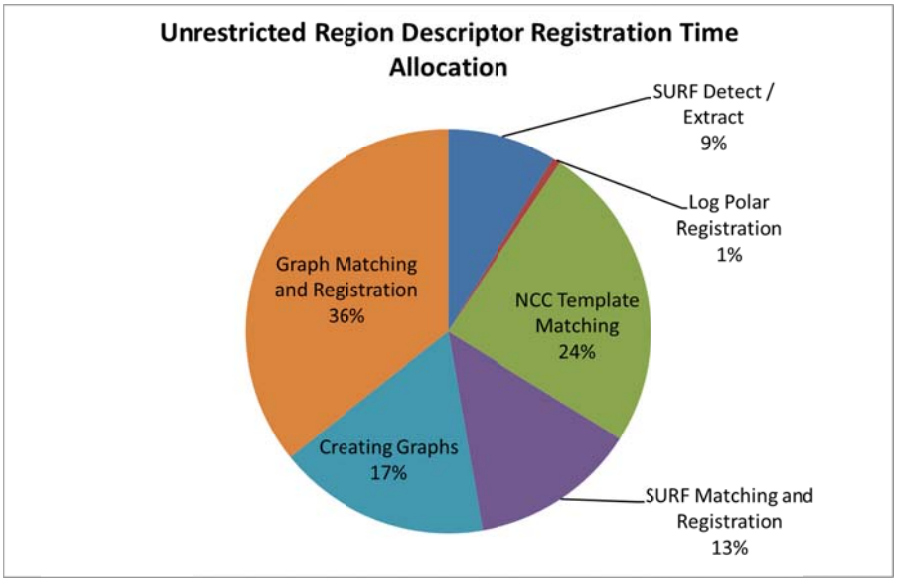
| | | Phase | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | SURF FP Detect / Extract (s) | Log-polar Coarse Registration (s) | Clustering and Graph Creation (s) | NCC Template Matching (s) | Limited Window Graph Matching (s) | SURF Matching (s) | Unrestricted Graph Matching (s) |
| Stage | Limited Window Region Descriptor | 1.570326087 | 0.06129268 | 1.830347826 | 0 | 2.365152174 | 0 | 0 |
| | SURF + NCC | 1.484268293 | 0.07013 | 0 | 1.979432 | 0 | 2.166512 | 0 |
| | Unrestricted Region Desriptor | 1.3783 | 0.68812748 | 1.5828 | 1.676533 | 0 | 2.1109 | 3.1851 |




a. Time allocation for the limited window registration approach




b. Time allocation for the SURF registration scheme




c. Time allocation for the unrestricted region descriptor registration

Fig. A.2 Time allocations for each stage of the proposed method

Many factors contribute to the overall execution time. Factors such as image resolution and image composition directly affect the number of feature points identified. Furthermore, the choice of feature point detector will greatly influence the latency incurred by the proposed algorithm. For instance, the Harris corner detector may identify hundreds of corners within an urban scene whereas the same detector may not produce many feature points from an aerial image of a natural scene. The number of detected graph-based region descriptors is linearly affected by the number of feature points, as shown in Fig. A.3.

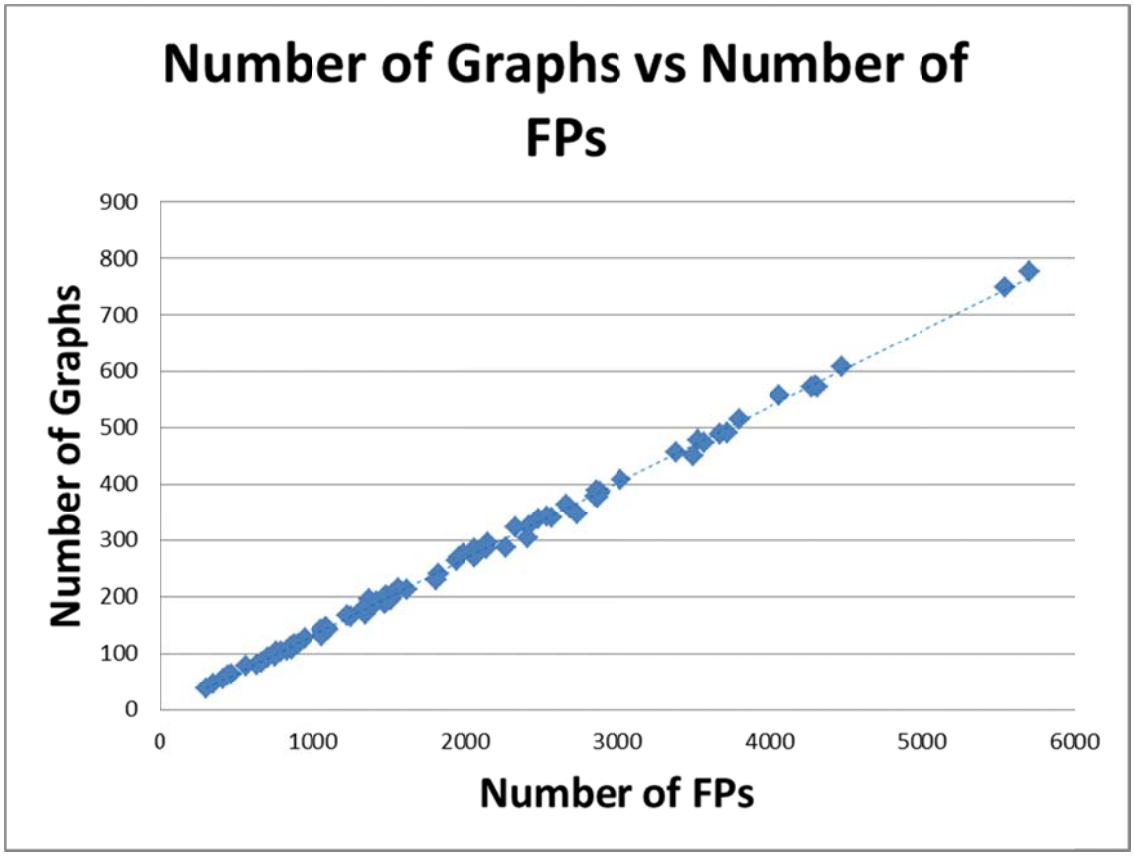



Fig. A.3 Relationship between the number of detected region descriptor graphs and the number of initial feature points detected

The relationship between the number of detected feature points and the execution times for each phase is given in Fig. A.4. The data plotted in Fig. A.4 show that the limited window region descriptor registration method performs comparably to the SURF matching using a coarse template matching scheme, however as would be expected, the unrestricted region descriptor registration approach incurs the most latency due to the approach requiring the previous two methods to fail before being attempted.

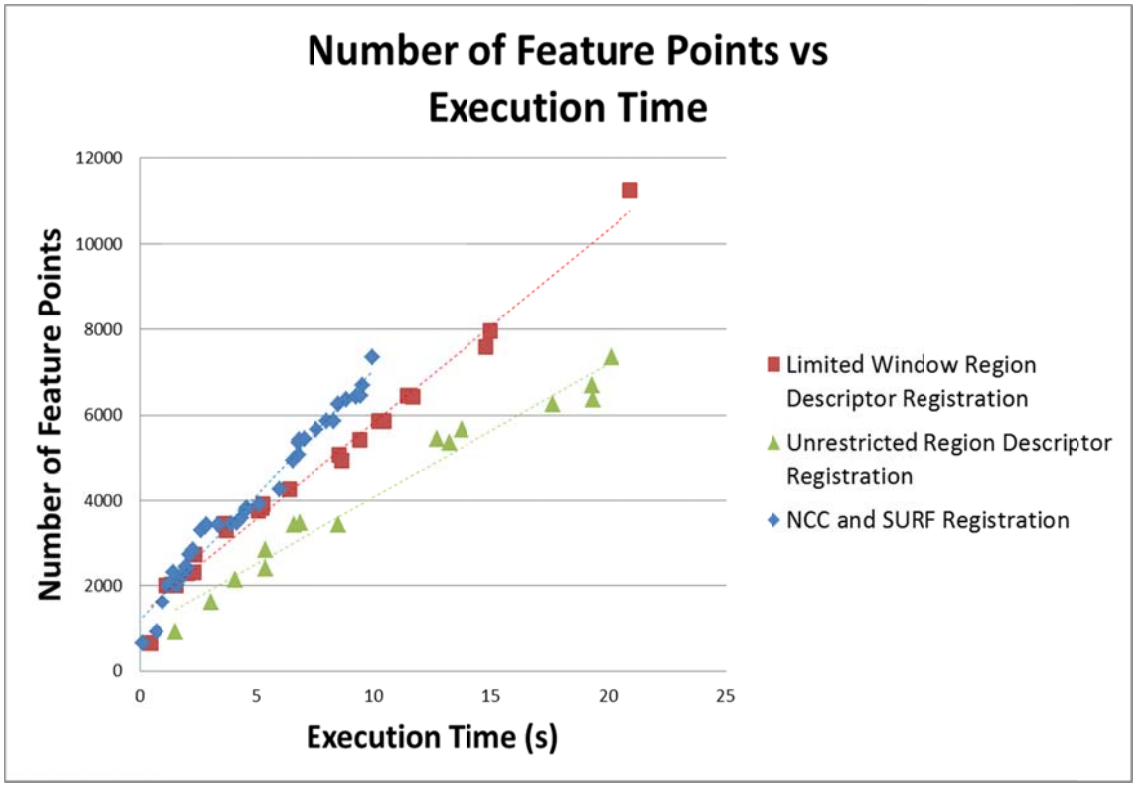



Fig. A.4 Number of feature points related to execution time for each proposed stage

As previously outlined, the major tasks that make up the proposed method can all benefit from the utilization of a GPU. Recent examples have been provided to suggest massive speedups. Table A.4 summarizes the reported speedups for each task.

Table A.4 Speedups using GPU architecture

| Task | Speedup |
|---|---|
| SURF | 46 |
| Clustering | 40 |
| Log Polar Registration | 3.43 |
| Graph Matching | 32 |
| Template Matching | 137.9 |

After applying the speedups to the average times listed in Table A.3, we derive estimated execution times for each phase and provide the summary in Table A.5. The average overall execution times are estimated to be 174ms for the limited window registration method, 114ms for the SURF registration with coarse template matching, and 428ms for the unrestricted region descriptor registration.

Table A.5 Estimated average execution times for the phases of each stage of the proposed method while utilizing a GPU

| | | Phase | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | SURF FP Detect / Extract (s) | Log-polar Coarse Registration (s) | Clustering and Graph Creation (s) | NCC Template Matching (s) | Limited Window Graph Matching (s) | SURF Matching (s) | Unrestricted Graph Matching (s) |
| Stage | Limited Window Region Descriptor | 0.034138 | 0.017869587 | 0.0457587 | 0 | 0.073911 | 0 | 0 |
| | SURF + NCC | 0.032267 | 0.020446064 | 0 | 0.014354 | 0 | 0.047098 | 0 |
| | Unrestricted Region Desriptor | 0.029963 | 0.200620258 | 0.03957 | 0.012158 | 0 | 0.045889 | 0.099534375 |

## A.9 Summary

In this chapter, a thorough review of computational complexity is provided. The major tasks of the proposed algorithm are analyzed while recent literature is cited to provide example systems that achieve great speedups which can be directly applied to the proposed methods.

VITA

# VITA

Francis Bowen earned a Bachelor of Science in Electrical and Computer Engineering at Indiana University-Purdue University Indianapolis in 2005 where he was awarded the Dunapace Senior Design Award for his RFID medical shelf. In 2007 he was granted a Master of Science from the Electrical and Computer Engineering Department at Indiana University-Purdue University Indianapolis. His research focus during this period was Field Programmable Gate Arrays and mitigating single event upsets in the presence of radiation.

Francis, an Indiana University Fellowship recipient, started the Ph.D. program at Purdue University in 2008 where his focus became image processing. His major advisors are Dr. Eliza Y. Du and Dr. Jianghai Hu. His research interests include image processing, computer vision, and pattern recognition. While working on his doctoral thesis, Francis has collaborated on image processing projects for the National Science Foundation and has published papers in top conferences. In 2012, Francis won the Best Student Paper Award at the IEEE Systems, Man, and Cybernetics conference in Seoul, Korea.